Evolution of genetic diversity at two hypervariable markers: a microsatellite and the MHC

Jarl Andreas Anmarkrud

Dissertation presented for the degree of Philosophia Doctor

Natural History Museum

Faculty of Mathematics and Natural Sciences

University of Oslo

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 1053

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Supervisor

Jan T. Lifjeld

Professor

Natural History Museum

University of Oslo

Norway

Co-supervisor

Lutz Bachmann

Professor

Natural History Museum

University of Oslo

Norway

Co-supervisor

Arild Johnsen

Associate professor

Natural History Museum

University of Oslo

Norway

Co-supervisor

Oddmund Kleven

Post Doc

Natural History Museum

University of Oslo

Norway

Adjudicating committee

1st opponent

Craig Primmer

Professor

Department of Biology

University of Turku

Finland

2nd opponent

Helena Westerdahl

Researcher

Department of Biology

Lund University

Sweden

Administrative leader

Tor A. Bakke

Professor

Natural History Museum

University of Oslo

Norway

Acknowledgements

First of all I would like to thank my supervisor Jan. It has been a privilege to work with you and have you as my supervisor. Your enthusiasm, knowledge and support have been a great inspiration, and I join ranks with many others in appointing you the best supervisor in the country.

I am also grateful for the support and advice I received from my co-supervisors Arild, Lutz, and Oddmund. Arild, thank you for always being helpful and for being an optimistic, enthusiastic and caring colleague. Lutz, I am impressed by your intellect and your careful sense for details. Thank you for your always providing time and insight when I have asked for your help. Oddmund, your efficiency and energy is a hard example to follow. Thank you for always helping me when I have been lost in the jungle of statistics, and for always being available to answer my questions.

Gunnhild, Lars Erik, Frode F., Terje, Melissah, Silje, Jostein, Trond, or "the bird group", deserve special thanks. I am lucky to have been a part of such a friendly, dynamic and supportive working environment, and I have really enjoyed the numerous discussions we have had during coffee-breaks and lunches. I have looked forward to going to work almost every morning, and that is a privilege! Eirik R. and Halvard, thanks for helping me with phylogenetic analyses and for making me laugh by continuously dropping funny comments. I am also grateful to all my colleagues at the Natural History Museum for making my period as a PhD student a memorable time.

The field work in Heimdalen and Røros was a novel experience for a lab geek like me, and it has been both a great experience and an excellent break from the lab bench. Heimdalen will forever be a place of blue skies and warm sun, just as it was throughout the three field seasons I spent there. Thanks to all that were involved and helped made this work a pleasant memory.

Several collaborators have been instrumental in helping me finish this thesis; Miguel Alcaide, thank you for fruitful and interesting discussions regarding genotyping of avian MHC and for kindly sharing your data. Jacob Höglund, Tanja Strand and Peter Halvarsson, thank you for introducing me to the RSCA technique and for kindly hosting me in Uppsala. Henrik Pärn, thank you for considerable help with statistical analysis. I am

also grateful to all my co-authors for providing samples and valuable comments to drafts of initial manuscripts.

Finally, I would like to thank my friends and family for their support, valuable breaks and for reminding me of the life outside academia. Special thanks go to my fiancé, Kristine; Thank you for your love and patience. This PhD would not have been completed without your endless support. And to Sofie: The happy face you provide every day when I come home from work has given me the energy to continue those days when my PhD felt like an infinite project.

Oslo, February 2011.

Contents

List of papers

The thesis is based on the following papers and will be referred to in the text by their Roman numerals given below:

I Anmarkrud JA, Kleven O, Bachmann L & Lifjeld JT. Microsatellite evolution: Mutations, sequence variation, and homoplasy in the avian microsatellite locus *HrU10*. 2008. *BMC Evolutionary Biology* **8**, 138

II Anmarkrud JA, Kleven O, Augustin J, Bentz KH, Blomquist D, Fernie KJ, Magrath MJL, Pärn H, Quinn JS, Robertson RJ, Szép T, Tarof S, Wagner RH & Lifjeld JT. Factors affecting germline mutations in a hypervariable microsatellite: a comparative analysis of six species of swallows (Aves: Hirundinidae). *Submitted manuscript.*

III Anmarkrud JA, Johnsen A, Bachmann L & Lifjeld JT. Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes. 2010. *Journal of Evolutionary Biology* **23**, 1206-1217.

IV Anmarkrud JA, Kleven O, Johnsen A & Lifjeld JT. MHC diversity is positively associated with promiscuity in passerine birds. *Submitted manuscript*.

Summary

Genetic variation is a prerequisite for evolution. The present thesis examines how genetic diversity arises and is maintained. The thesis can be divided into two subprojects, addressing neutral and functional genetic variation, respectively. In the first subproject, germline mutation patterns on the presumably non-functional microsatellite locus *HrU10* were investigated. By sequencing observed mutations uncovered by microsatellite length typing, a significant association between microsatellite length and number of interrupting motifs (sequence motifs deviating from the repeated core unit) was found. Moreover, a majority of the mutations revealed homoplasic alleles. This result may imply a general caution in population genetic studies where genotypes are based on allele sizes only. In a comparative analysis of *HrU10* mutations in six species of swallows, the following sex specific patterns in *HrU10* mutability were found: (i) Mutations in long alleles were typically attributed to male germlines and mutations in short alleles were most pronounced in female germlines. (ii) Mutations in the male germlines were significantly biased for expansions while no such effects were uncovered in female germlines. These results indicate fundamental dimorphism in male and female gametogenesis with respect to slippage mutations.

The aim of the second subproject was to study the polymorphism at genes of the Major Histocompatibility Complex (MHC). The gene products of the MHC are important for immune response and subsequently individual fitness. First, the polymorphism at exon 2 in bluethroat (*Luscinia svecica*) MHC class II genes was assessed. The results revealed at least eleven different loci, and the amino acids involved with antigen recognition were subjected to significant diversifying selection. MHC class IIB (MHCIIB) exon 2 sequences from the thrush nightingale (*L. luscinia*) were also analyzed for comparison. Phylogenetic analyses indicate an ancestral origin of some of the polymorphism uncovered at the MHCIIB exon 2. Secondly, a standardized multispecies genotyping strategy was developed to estimate MHC diversity at eight closely related passerine species with known levels of promiscuity. Interestingly, an association between MHCIIB exon 2 diversity and promiscuity levels was identified. This result may indicate that promiscuity in passerine birds is a mechanism creating MHC diversity, and consequently promoting offspring immunological quality.

General introduction

Modern systematics, taxonomy and evolutionary biology are to a major extent studied at a molecular level. To make such studies possible, particular genetic characteristics are fundamental for categorizing biological units. Specific genetic features make it possible to classify molecular traits present at an individual, population or species level. Understanding how such molecular traits, or *genetic diversity*, arises and is maintained, is one of the basic themes in the field of evolutionary biology

Every cell in every living organism has a *deoxyribonucleotide acid* (DNA) molecule, encompassing the genetic information needed to construct the required RNA molecules and proteins. In other words, this molecule holds the 'recipe' for each organism. The *nucleotide* is the vital chemical component of the DNA[1] and consists of three main components: (i) a phosphate, (ii) a deoxyribose sugar and (iii) one of the four nitrogenous bases adenine, guanine, cytosine or thymine). The arrangement of nucleotides is termed a *nucleotide sequence*. In general, DNA sequences can be categorized as *coding* or *non-coding*. The fraction of the DNA that covers the coding sequences hold the information required for expression of products encoded in the genome. More precisely, coding DNA is the fractions of the genome that can be transcribed to RNA and translated into protein molecules[2]. Such segments encoding functional proteins are commonly termed *genes*. Nevertheless, an universal accepted definition of the term 'gene' has not yet been established[3], almost 150 years after Georg Mendel introduced inherited characteristics as the first 'geneticist'[4].

A large majority of the genome does not encode proteins and is referred to as 'non-coding' DNA. That is, non-coding functional RNAs[5], regulatory elements (e.g. Hentze & Kühn[6]), introns[7], transposons and viral elements[8], pseudogenes[9], teleomeres[10] and tandemly repeated sequences, i.e. microsatellites[11]. Much of the non-coding DNA has no functional potential and has been designated as 'junk DNA'[12]. However, mounting evidence suggests that presumed 'junk DNA' may have important biological functions[13].

The present thesis can be split into two subprojects. The first focuses the subject of microsatellite evolution. In the second subproject, evolution of the genes of the major histocompatibility complex (MHC) was investigated. Microsatellites and MHC genes are representing non-coding and coding DNA, respectively. Accordingly, the mode of

evolution is acting differently on the two types of DNA, with respect to mutation rates, mutational mechanisms, evolutionary history and selective mechanisms.

Microsatellites

Microsatellites (also referred to as short tandem repeats or simple sequence repeats) are tandemly repeated nucleotide sequences no more than six base pairs long. They are frequently scattered throughout the genomes of all organisms analyzed[14], and comprise roughly 3 % of the human genome[15]. Such tandemly repeated sequences may be vital components in protein coding genes (e.g. Manto[16]) and important in several regulative processes[17,18]. However, most microsatellites are considered as non-coding DNA with no functional potential[19,20]. As a consequence, microsatellite evolution is expected to reflect neutral selection processes. Microsatellites also possess an extensive polymorphism by means of allele lengths and this allele size variation is manifested in different numbers of repeated microsatellite units.

Microsatellites mutate through the process of replication slippage[21]. Replication slippage involves a gain or loss of one or more tandemly repeated units during replication, compatible with a step-wise mutation model[22]. Another striking feature of microsatellites is the mutation rates, which on average is several orders of magnitude higher than non-repetetive DNA, [23]. Allele length by means of numbers of tandemly repeated units is commonly accepted as the most important factor promoting the mutation rate[21,24-34].

Microsatellites are presumably selectively neutral and relatively easy genotype. Because of these features, they have been the marker of choice in many applications, such as e.g. genetic structure of populations, parentage analyses and assessment of phylogenetic differentiation[11].

MHC genes

MHC genes encode glycoproteins that expose foreign infectious antigens on the cell surface and present these antigen to T-cells which in turn trigger a cascade of immune responses[35]. Accordingly, these gene products are important for immune function and individual fitness. There are three classes MHC genes. MHC class I genes are expressed in most somatic cells and they are associated with antigen recognition of intracellular material (e.g. viruses and cancer infected cells). Products of these genes present pathogenic substances to $CD8^+$ receptors in T-cytotoxic cells[36,37]. Extracellular infections, on the other hand, are associated with MHC class II genes that are expressed by 'specialized' antigen

presenting cells (B cells, macrophages, dendritic cells). Parasite resistance is subsequently connected to MHC class II genes, in which MHC Class II peptides present pathogenic substances to $CD4^+$ receptors in T-helper cells[38,39]. The function of MHC class III genes are very different from MHC class I and MHC class II genes and not relevant for this thesis.

The structural conformation of mature antigen presenting MHC class I and class II peptides is relatively similar, both put together as a heterodimer with four extracellular domains[40,41]. Class I molecules contain a heavy (44 kDa) α chain and a light (12 kDa) β_2-microglobulin chain. The α chain is divided into three domains where the α_3 C-terminal domain is attached to the cell membrane. The α_1 and α_2 domain form a platform consisting of eight antiparallel β-sheets topped by two α-helixes creating a groove between the helixes[40]. Opposite to Class I peptides, the α and β chain in the peptides of the Class II heterodimer are more similar in weight (34 kDa and 28 kDA). Both chains have two domains where the α_2 and β_2 domain bind to the cell membrane (Fig. 1) and the α_1 and β_1 domain create a groove with β-sheets topped by two α-helixes, similar as the Class I peptides[41,42]. Amino acids surfacing this groove provide binding sites for processed foreign antigens. Most of the amino acids in the peptide binding region (PBR) are located in exon 3 of class I genes and exon 2 of the Class II B genes.

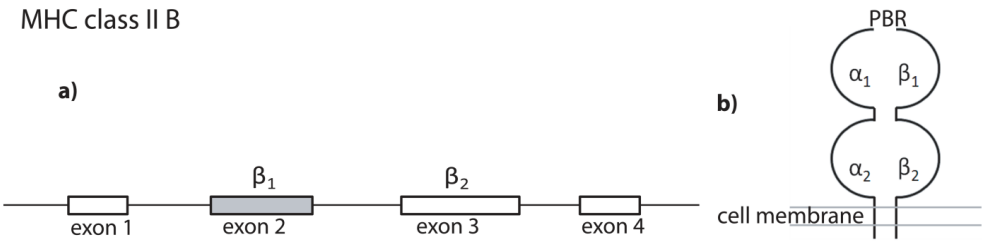

Figure 1. Organization and structure of MHC class II B. (a) Exon 2 (filled box) encode the β_1 domain. Most of the amino acids in the PBR are encoded by this domain. The model of the mature peptide (b) illustrates the four domains in the heterodimer. The PBR are located as a groove between the α_1 and β_1 domain.

MHC genes are extremely variable[43,44]. This extensive polymorphism is due to strong diversifying selection pressure. The amino acids involved with antigen recognition, the PBR, are the most important for immune function and most of the variation observed in MHC genes can be attributed to this region[45,46]. MHC polymorphism is mediated by

parasite-host co-evolution[47,48]. Furthermore, in order to increase offspring immunological competence for parasite exposure, the genes of the MHC are expected to be subject to sexual selection[49]. Nevertheless, the selective mechanisms creating the huge variation have been extensively debated[47,50-52].

Mating behavior and evolution of genetic diversity in passerine birds

Even though the majority of passerine birds are socially monogamous, most species frequently copulate outside the pair bond[53]. In fact, approximately 90 % of passerine birds examined produce extrapair offspring[53]. Increased number of sired offspring explains male motivation for such behavior. However, females produce only a limited number of eggs, and may, consequently, only influence quality of offspring, not quantity. Accordingly, the female benefits of extrapair paternity (EPP) have been one of the enigmas for evolutionary ecologists[53-56].

A role for sexual selection to increase mutation rate and thus boost the evolution of genetic diversity has been suggested both theoretically and empirically[57-59]. Elevated mutation rates in species attributed to intense sperm competition may inhibit the fixation process expected in sexually selected characters, and maintain variation in these features[58]. Accordingly, promoted mutation rates may be a solution to the 'lek paradox', which refers to the dilemma of how variation can be maintained in sexually selected characters[60]. Interestingly, a link between mutation rates and EPP levels has been documented in tandemly repeated DNA[57,61], although not in microsatellites.

In passerines, several studies have documented associations between EPP and immunological quality in offspring, congruent to the compatible genes hypothesis[62-64]. Likewise, gene products of the MHC are important for immune response. One may also expect a correlation between 'allelic richness' at the MHC and capabilities for parasite resistance. Consequently, female mate choice for compatible genes may be important to promote offspring parasite resistance and fitness. Because passerine species vary remarkably in promiscuity levels[53], and display a variable MHC[65], the MHC in song birds is a promising system to investigate potential links between promiscuity and MHC diversity.

Aims of the study

The main objective in the present PhD project has been to investigate the mechanism creating genetic diversity, using selected passerine birds as model organisms. This aim was approached in two ways. First, by examining the evolution of neutral microsatellite DNA and, secondly, by studying genes of the MHC, coding DNA fragments under strong selective pressure.

Project 1:

Microsatellite evolution- case studies on the avian pentanucleotide HrU10

Gaining empirical mutation data is difficult because mutations occur at low frequency and parentage information from non-model organisms is demanding to obtain. Consequently, empirical studies regarding microsatellite mutations in natural populations are relatively limited, and mostly based on fragment length analyses only. Such analyses, without sequence data, may not uncover essential information i.e. substitutions (replacement of a single nucleotide), patterns of interrupting motifs (sequence motifs deviating from the repeated core unit) and homoplasy (alleles similar in state but not in descent). Although several factors are important for inducing microsatellite mutation rates, a large body of evidence suggest allele size to affect microsatellite stability, with increased mutation rates in longer alleles[23]. Consequently, one may also assume increased probability for introduction of interrupting motifs in longer microsatellite motifs compared with a di-nucleotide. Longer motifs will require a larger spatial conformation during loop formation in a slippage event. If the loop on the nascent or template strand not includes a complete pentanucleotide unit during a slippage event, somewhat misaligned nucleotides might occur and thereby introduce interrupting motifs.

Based on a large pedigree study in barn swallow (*Hirundo rustica*)[66], several germline mutations consistent with a process of replication slippage were uncovered in the presumably neutral pentanucleotide microsatellite *HrU10*. This marker has also previously been shown to be very variable with high mutation rates[28].

Because the *HrU10* locus is variable and hypermutable, this microsatellite is a good candidate marker to test a hypothesis of elevated mutation rates and increased probability for introduction of interrupting motifs in longer microsatellite alleles. By sequence

analyses of both parent and mutant offspring in alleles identified with mutations in the germline, one may investigate the formation and magnitude of homoplasic *HrU10* alleles (Paper I).

Furthermore, one aim of the thesis was to empirically examine whether microsatellite mutation rates could be linked to mate choice and sexual selection. Thus, in Paper II we screened for *HrU10* mutations in specimens from six species of Hirundinidae with known paternity data and EPP levels; Barn swallow[66], fairy martin (*Hirundo ariel*) (M. J. L. Magrath, unpublished), house martin (*Delichon* urbica)[67], purple martin (*Progne subis*) (S. Tarof, unpublished), sand martin (*Riparia riparia*)[68] and tree swallow[69]. By applying this approach, we obtained mutation data from the same genetic marker from natural population in six passerines, providing an opportunity to comparatively investigate mutational dynamics of *HrU10* between six species in the same family (Paper II).

Project 2:

Evolution of MHC diversity in passerine birds

Many groups of birds have a relatively simple organization of the MHC[70-74], also referred to as the 'minimal essential MHC'[75]. A growing number of evidences suggest, on the other hand, that the 'minimal essential MHC' is not representative for the most species rich group of birds; Passeriformes[65,76,77]. In fact, recent published results indicates the passerine MHC to be the most variable discovered to date[77]. Such diversity may reflect local adaptation due to parasite-host arms race[78], or being the retention of allelic lineages in a longer evolutionary time perspective, predating recent speciation processes. The latter also referred to as trans-species- or *ancestral polymorphism*[79].

In Paper III we aimed to describe, using the bluethroat as a model (Fig. 2), polymorphism and evolution of exon 2 of the MHC class II (MHCIIB) genes. By reconstructing phylogenetic history of MHCIIB exon 2 in the bluethroat and the congeneric species, thrush nightingale (*Luscinia luscinia*), we assessed the extent of ancestral polymorphism in these species. Finally, by comparing substitution patterns (non-synonymous (d_N) vs. synomynous (d_S) substitutions) in the PBR compared to the non-PBR, we investigated selective signatures in the bluethroat MHCIIB exon 2 genes.

Figure 2. Male bluethroat captured in Heimdalen, Norway. Photo: Jarl A. Anmarkrud.

Sexual selection is assumed to be an important mechanism creating MHC diversity, also in humans[49,80-84]. The issue of MHC based mating preferences in humans is controversial and has been extensively debated[85-87]. Nevertheless, during the last 15 years, a link between mating preferences and MHC has been uncovered in all groups of vertebrates; fish[88-90], amphibians[91], reptiles[92], birds[93-95] and mammals[82,85]. Even though studies on mating behavior is frequent in passerines[53], studies concerning mate choice in association to genes of the MHC are limited. Such associations is challenging to study empirically due to the extensive MHC variation observed in passerines, hampering MHC genotyping protocols.

In Paper IV we developed a standardized study design to genotype MHCIIB exon 2 in multiple passerine species. Thus, by employing identical experimental conditions to all the species, we provided comparable genetic diversity estimates. By using these diversity estimates, we examined potential links between promiscuity and MHC diversity, by using two independent proxies for promiscuity: (a) between male variation in sperm lengths and (b) proportion of extrapair offspring.

General methods

Except for one individual bluethroat captured in Heimdalen, Norway, for the purpose of RNA extraction, the study objects used in this thesis were obtained from the DNA/tissue collection at the Natural History Museum in Oslo or provided by colleagues. DNA was isolated from blood or tissue using standard DNA extraction kits and amplified employing standard polymerase chain reactions (PCR)[96].

Microsatellite genotyping

Amplified *HrU10* and other amplified microsatellite fragments were genotyped by capillary electrophoresis (CE) utilizing an ABI 3100/3130xl (Applied Biosystems). The principle for this technique is to separate DNA fragments by size when operated through a capillary matrix. The fragments labeled with fluorescence can subsequently be identified by a laser-detector after electrophoresis. If pedigree is known, mutations can be detected by direct observation of parent-offspring allele transmission (Fig. 3).

MHC genotyping

Targeting multiple loci by PCR is not straightforward. Several aspects must be taken into consideration when employing primers annealing to multiple MHC targets, both concerning formation of PCR artifacts and genotyping protocol[97].

Formation of sequence chimeras is a potential drawback when amplifying multiple loci by PCR. Sequence chimeras are PCR mediated recombination, where incompletely elongated amplicons function as primers by hybridizing to a wrong template[98]. Such sequence chimeras may be a major pitfall when genotyping genes of the MHC[99]. However, the proportion of chimera artifacts can be reduced using simple steps, i.e. reduce number of cycles and increase elongation time in the PCR[100,101].

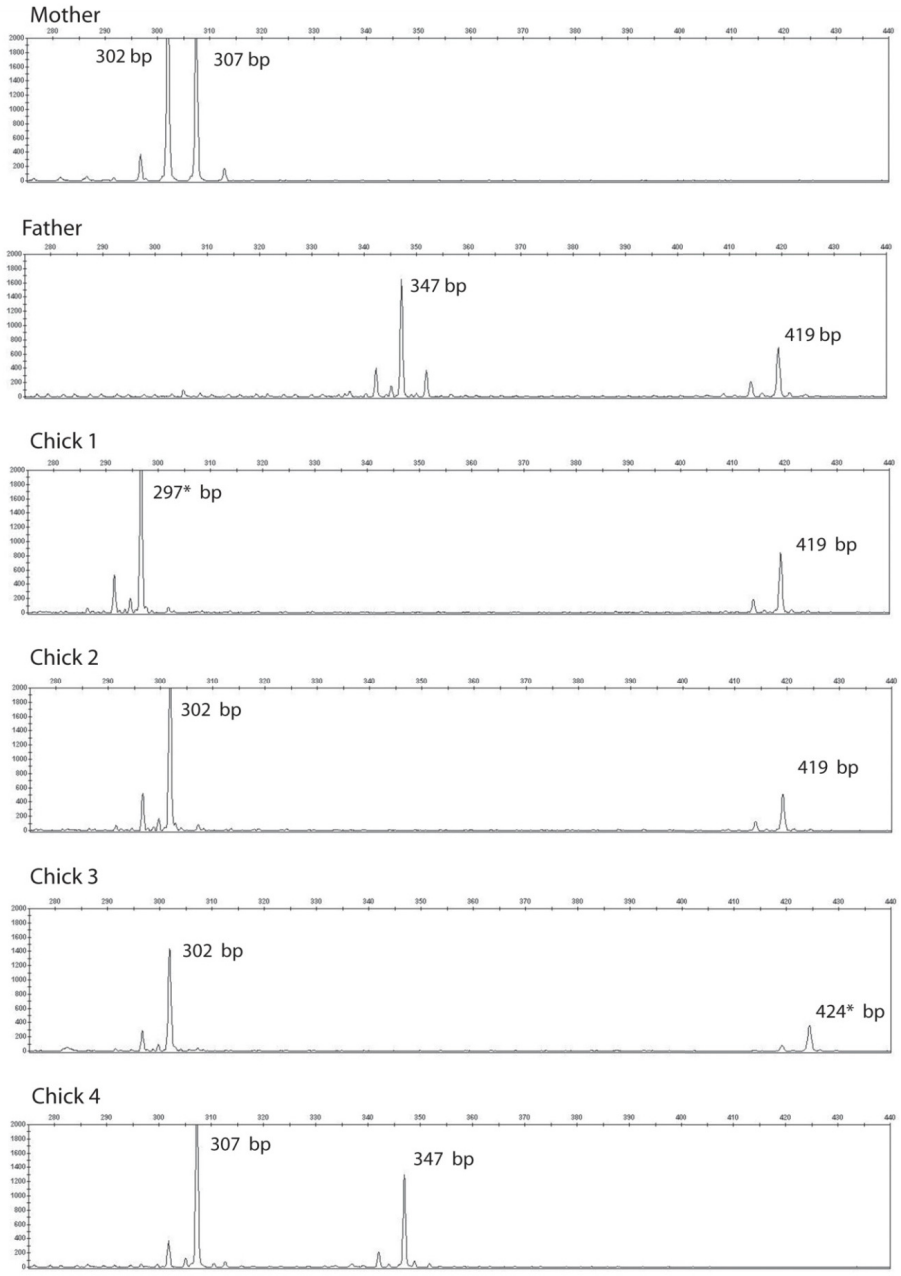

Figure 3. Screening for *HrU10* mutations in known pedigrees. Detected *HrU10* alleles are shown as tall peaks and allele sizes are given in the electropherogram. Slippage mutations were confirmed if alleles were incompatible with Mendelian inheritance. That is for the case of *HrU10*, a gain or loss of one (or more) pentanucleotide unit (asterisks). These particular genotypes are from a family of purple martins.

Initial assessments of MHC sequences are usually obtained by cloning and sequencing of PCR products, and this strategy was employed in Paper III. MHCIIB exon 2 amplicons from both genomic DNA and cDNA (reverse-transcribed RNA, for verification of transcription) was ligated into a vector and subsequently transformed into a bacteria strain which were plated on a growth medium containing an appropriate selection marker. Individual bacteria colonies harboring homogenous inserts originated from the same DNA molecule were then used as template in PCRs. This technique has the advantage of obtaining nucleotide sequence information from single PCR products from heterogenic amplicons. However, cloning is resource demanding and not well suited for large scale analyses due to the vast amount of clones needed to be sequenced for accurate genotyping when multiple loci are targeted.

Large-scale genotyping of MHC genes are most commonly performed by PCR based electrophoresis techniques; e.g. denaturing gradient gel electrophoresis (DGGE)[102], single strand conformational polymorphism (SSCP)[103] or reference strand conformational analysis (RSCA)[104]. The last two, SSCP and RSCA, may be performed in CE instruments such as ABI genetic analyzers. The principle in these methods is to separate fragment by CE based on secondary structure. DNA molecules with different sequence will fold differently, and allele variation can be manifested in specific detectable conformations. One of the initial intentions with the PhD-project was to develop a protocol for large-scale genotyping of bluethroat MHCIIB exon 2 genes. By this approach, we aimed to screen for an MHC dependent genetic component, corresponding to a previously uncovered immune function in this species[64]. This immune function was enhanced in extrapair young, and thus in accordance with a 'compatible genes' hypothesis. A substantial effort was put into the development of a CE-SSCP and CE-RSCA MHCIIB exon 2 genotyping protocol. During the primer optimization process, the primer combinations targeted either specific alleles or a large number of loci. Furthermore, when testing a primer combination optimized for amplification of a restricted number of classical polymorphic MHCIIB exon 2 loci in the European pied flycatcher (*Ficedula hypoleuca*) for CE-SSCP analysis (See Canal et al.[105] for details), 20-30 peaks were uncovered in the following electropherogram obtained from bluethroat amplicons (unpublished results). Unfortunately, the extensive variation by means of duplicated loci made it impossible to verify unique alleles in the obtained electropherograms (data not shown).

Due to the low success in the large scale SSCP approach of MHC genotyping bluethroat MHCIIB exon 2 genes, an alternative strategy was applied to investigate potential links between MHC diversity and female promiscuity (Paper IV).

Summary of papers

Paper I

Microsatellite evolution: Mutations, sequence variation, and homoplasy in the avian microsatellite locus HrU10

In Paper I, we tested the hypothesis of increased probability for introductions of mutations and interrupting motifs with increased microsatellite length in the avian pentanucleotide marker *HrU10*. Indeed, long *HrU10* alleles were both more mutable and had more interrupting motifs in barn swallows and tree swallows. Nevertheless, no correlation between longest stretch of perfect tandem repeats and total allele size was observed. This result indicates a threshold level for maximum number of perfect repeats before an IM is introduced in this microsatellite. The sequence data also revealed that a majority of the mutations created new DNA fragments, size homoplasic to present alleles.

Paper II

Factors affecting germline mutations in a hypervariable microsatellite: a comparative analysis of six species of swallows (Aves: Hirundinidae)

By applying a comparative approach, we aimed to screen for *HrU10* mutagenesis between species. Amplicons from this genetic marker were generated from six species of Hirundinidae. A substantial amount (n = 100) of *HrU10* slippage mutations were uncovered. The different mutation rates could not be attributed to mate choice assessed as EPP. Mutation rates differed between the species (range 0.5 % - 10.6 %) and mutations were most frequent in female germlines. Although considerably different mutation rates were uncovered between the species, this effect was negligible when controlling for allele size. In other words, allele size and sex were essential factors inducing *HrU10* slippage mutations. Moreover, mutations in short alleles were mostly maternally transmitted while mutations in long alleles were mostly in male germlines. Such sex specific heterogeneity was also documented in mutation directionality. Male mutations were significantly biased for expansions, while no such effects were uncovered in the female mutations. Our results provide evidence for distinct sex-specific slippage dynamics during meiosis, thus implying sex specific mechanisms creating microsatellite diversity.

Paper III

Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes

In Paper III, we examined variation, evolutionary history and selective signatures in exon 2 of bluethroat MHCIIB genes, using samples from three distinct subspecies; *L. s. svecica, L. s. cyanecula* and *L. s. azuricollis*. Our results revealed a substantial polymorphism in these genes, both by means of number of different alleles and number of loci. Combining sequence information obtained from genomic DNA and cDNA, a minimum of eleven presumably functional MHCIIB exon 2 loci was uncovered. Phylogenetic analyses, including MHCIIB exon 2 sequences from the thrush nightingale, revealed a gene tree with shared clades between the two species. The shared clades were poorly supported and the overlap in allelic repertoire indicates an ancestral origin of the polymorphism, predating the phylogenetic split between the species. Furthermore, dN/dS values were significantly higher in the PBR compared to the non-PBR, suggesting strong diversifying selection in the amino acids involved with antigen recognition.

Paper IV

MHC diversity is positively associated with promiscuity in passerine birds

By developing a standardized genotyping protocol, we aimed to provide comparable diversity estimates for MHCIIB exon 2 in a multispecies approach. These estimates were applied to test for a potential association between promiscuity and MHC polymorphism. Interestingly, the MHC diversity was significantly correlated to both proxies for intensity of promiscuity (between male variation in total sperm lengths and EPP) in the PBR of MHCIIB exon 2. In other words, our results indicates that MHCIIB exon 2 polymorphism may by promoted by mate choice preferences for increased MHC diversity in highly promiscuous passerines. Even though the sample size is relatively restricted, the relationship between EPP and MHC diversity was striking. Thus, these results may indicate that female promiscuity is an important mechanism enhancing diversity at the MHC genes in passerine birds.

General discussion

It is essential for every organism to continuously adapt to a dynamic environment. Consequently, genetic variation is fundamental in nature. Without genetic changes, there would be no new alleles or new genes and, accordingly, no evolution. In other words: 'mutations are the raw material of evolution'[106].

The present thesis aims at investigating the mechanisms creating genetic variation, approaching evolution of two distinct variants of DNA sequences: microsatellites (Paper I, Paper II) and MHC genes, respectively (Paper III, Paper IV).

In project 1, we have identified a positive association between *HrU10* allele size and interrupting motifs. Furthermore, we discovered a sex specific pathway in *HrU10* mutability, which were affected by allele size across species borders. In project 2, we documented large MHCIIB exon 2 variation in the bluethroat in terms of nucleotide diversity and duplicated loci. This variation is likely of ancient origin maintained by strong diversifying selection pressure. Interestingly, in a survey of MHC diversity in eight related passerine species, we found indications that the MHCIIB exon 2 polymorphism may be promoted by female promiscuity.

***Project 1*: Microsatellite evolution**

Both Paper I and Paper II were case studies, examining the mode of evolution in the hypervariable pentanucleotide *HrU10*[107]. *HrU10* is mapped to chromosome 18 in the chicken (*Gallus gallus*) genome[108], and is expected to be selectively neutral. As a selectively neutral marker, *HrU10* development is expected to reflect neutral genome wide microsatellite evolution. Even though *HrU10* mutagenesis may reflect genome wide microsatellite evolution, the evolutionary rate in terms of mutation rate per generation is far beyond average for this marker. In fact, a mutation rate of 10.6 % in sand martins is the highest ever reported for a microsatellite (Paper II).

HrU10 mutation patterns have also previously been studied, in European barn swallow populations[27,28]. These studies found a correlation between slippage rate and allele size, and a bias for maternally transmitted mutations. These results were confirmed by Paper I. In addition, we obtained sequence information from observed germline mutation events. The sequence analysis revealed, as expected, replication slippage as the most likely mutation mechanism. Furthermore, in Paper I we verified a significantly positive

association between allele size and introduction of interrupting motifs. No correlation between longest tract of perfectly repeated pentanucleotide units and number if interrupting motifs was observed. Such relatively constant maximal length of stretches of perfect repeats may indicate a threshold level for stable arrays of perfect repeated microsatellite units. This result may be explained by a selection pressure to retain stable secondary structures[109], or due to increased probability to introduce incomplete slippage in long alleles.

Another important result in Paper I was the frequent formations of size homoplasic alleles. More than 50 % of the sequenced *HrU10* mutations resulted in alleles size homoplasic to other sequenced alleles in the dataset. A large fraction (23 %) of these alleles was identical in both size and sequence. Such homoplasy is only detectable through sequencing of observed mutations and has to our knowledge not previously been empirically documented. Several theoretical models have been put forward to predict microsatellite evolution[22,110-112], and the issue of size homoplasy has been approached theoretically based on these mutation models[113]. However, the mutational dynamics of *HrU10* can not be applied to these models (see Paper I for details), a result illustrating the challenges of incorporate evolution into a theoretical framework.

In Paper II, we expanded the screening process of *HrU10* mutations to include six species of Hirundinidae and almost four thousand meiotic events. Among these meiotic events, 100 *HrU10* mutations were observed and the mutation rates ranged from 0.5 % (house martin) to 10.6 % (sand martin). Hence, this is to our knowledge the most comprehensive microsatellite mutation study performed on wild bird populations, both with respect to number of mutations, number of meiotic events for one microsatellite locus and number of species with data from the same hypermutable microsatellite marker.

The initial motivation was to test a hypothesis of elevated mutation rates in species subjected to intense sexual selection or high risk of sperm competition (using EPP as index for sperm competition). Directional selection among the choosy sex (females) may lead to fixation of sexually selected characters. However, considerable variation in fitness is likely to occur[114]. Accordingly, the total mutational load might be promoted in species subject to intense sexual selection[57,59,61] and potentially be a solution to the 'lek paradox'[60].

No relationship between mutation rates and proportion of EPP was uncovered among the swallow species. Taken into consideration that we analyzed one locus only, our result was not surprising. Especially when keeping in mind the fact that microsatellite slippage mutations are affected by local genomic effects (e.g. allele size). Møller and

Cuervo[57,61] documented a relationship between minisatellite mutation rates and EPP. However, the different mutation mechanisms in minisatellites and microsatellites might explain the observed inconsistence in mutation patterns between these two types of repeated DNA sequences[21,115].

A substantial different *HrU10* mutation rate was documented between the six species of swallows (Paper II). Moreover, a significant over all bias for mutation transmission in female germlines was uncovered. A result consistent with previous studies (Paper I)[28]. However, a tendency for an interaction of allele size and sex was revealed. That is, long alleles mutated in male germlines and short alleles mutated in female germlines. Furthermore, directionality of mutations was significantly distinct between the two sexes. Most male mutations increased in size, while no such effect was uncovered among the female mutations, a result also documented (although not explained) for another avian microsatellite, the *HrU9* locus[116].

We hypothesize this sex specific heterogeneity to be explained by the sex linked molecular mechanism during gametogenesis. The enzymatic machinery ensuring correct duplication of the genome during each cell cycle, the replication licensing system[117], has distinct sex-based pathways[118]. Striking dimorphism in expression patterns of vital replication licensing proteins has been documented in oogenesis and spermatogenesis, respectively[119]. Due to more cell divisions in male than female germlines, one may expect a bias for paternally transmitted mutations, because most mutations occur as replication errors. This issue have been thoroughly studied and discussed[120-128] since Haldane introduced the subject of sex dimorphic mutations rates[129]. Such male biased mutation rate has also been documented in bird species based on substitution patterns in sex chromosomes[130,131]. A male biased mutation rate has also been uncovered among multilocus examinations of human microsatellites[26,132,133]. Even though a tendency for male biased mutation was documented at the *HrU9* locus in barn swallows[116], significant sex effects in microsatellite mutability in non-mammals have been female biased[24,28,29,134,135]. It must be emphasized that the combined number of loci in these studies is seven. Accordingly, this issue needs to be examined more comprehensively in a genome wide approach.

Project 2: Evolution of MHC diversity in passerine birds

Substantial polymorphism was documented in exon 2 of bluethroat MHC class II genes. By combining genomic DNA and cDNA in one individual, we uncovered a minimum of

11 functional MHCIIB exon 2 loci. MHC polymorphism, in terms of numbers of loci, is assumed to be relatively restricted in non-passerine birds; Charadriiformes[71], Galliformes[136], Sphenisciformes[70,137], Falconiformes[138], Stringiformes[139] and Psittaciformes[72]. Compared to other groups of birds, the most species rich group, passeriformes, has extensive polymorphic MHC genes by means of multi-copied loci[65,76,77]. Recently, Bollmer et al.[77] hypothesized at least 20 MHCIIB exon 2 loci in the common yellowthroat (*Geothlypis tricas*).

The huge variation has been hypothesized to originate from a parasite-host arms race maintaining MHC variation[78]. Such parasite-host arms race may select for high MHC variation over a relatively short time perspective, and local adaptation to specific parasitic spectrums should be expected. Hence, local adaptation attributed to genes of avian MHC has been documented in migratory birds with restricted gene flow[71,138]. In Paper III, we analyzed samples from distinct subspecies of the bluethroat, both with respect to geographical origin and habitat preferences. However, no geographic structure in MHCIIB exon 2 alleles was identified. The bluethroat MHCIIB exon 2 alleles from the different subspecies showed, on the contrary, a striking overlap in allelic repertoire. Moreover, several clades of MHCIIB exon 2 sequences from the congeneric thrush nightingale clustered together with bluethroat MHCIIB exon 2 sequences. This result indicates an ancestral origin of the observed MHCIIB exon 2 variation, predating speciation and intra-specific differentiation processes[79]. Although ancestral polymorphism in MHCIIB have mostly been reported within mammals, a growing number of studies have discovered such signatures among avian species, e.g. raptors[140], owls[141], passerines[142-146] and penguins[137]. Nevertheless, potential locus specific geographical structure in the bluethroat MHCIIB genes may be masked due to the amplification of multiple loci.

We identified a minimum of eleven MHCIIB exon 2 loci in one bluethroat individual. Such polymorphism may have originated by gene duplication and evolved according to the concept of concerted evolution. Accordingly, the gene duplications and recombination events must have influenced the evolution of bluethroat MHC. Recombination has been hypothesized as an important mechanism shaping MHC polymorphism[147,148], also within birds[136,149-152]. Recently, Balakrishnan et al.[76] characterized the MHC in zebra finches (*Taeniopygia guttata*), the only passerine species with the complete genome sequenced[153]. Their results identified gene duplications and MHC genes located in several chromosomes. Our results (Paper III), together with Bollmer et al.[77] and Balakrishnan et al.[76], highlight the complex organization associated with

passerine MHC. This complex MHC organization is likely to have originated through ancient gene duplications, where the polymorphism is maintained through strong balancing or diversifying selection, and the MHC loci evolve in accordance with the birth-and-death process of multiple gene evolution[154,155].

In Paper IV, the processes creating diversity at the MHCIIB genes were investigated. When employing a comparable experimental design to assess different diversity estimates, the eight analyzed species revealed different levels of MHCIIB exon 2 polymorphism and strong diversifying selection in the PBR. The MHCIIB exon 2 variation could not be attributed to the different number of sequences obtained from the respective study organisms. On the other hand, the MHC diversity was significantly associated with not only the between male variation on total sperm length values from the particular study populations, but also known EPP levels.

The results presented in Paper IV indicate that promiscuity may be a mating strategy creating MHC diversity and, thus, a mechanism to maintain parasite resistance in a parasite host co-evolutionary arms race An immunological benefit in offspring sired by extrapair males has been documented in passerine species (bluethroat[62,64], common yellowthroat[63]). Accordingly, species with high promiscuity levels might have a relatively diverse MHC, which subsequently may give resistance to a larger spectrum of pathogenic substances than species with relatively low polymorphic MHC. Even though the results presented in Paper IV may suggest promiscuity to be a mechanism enhancing MHC diversity, we can not verify the causality of the correlation in this comparative approach.

MHC-dependent mate choice has been described in passerines (savannah sparrows (*Passerculus sandwichensis*)[156], Seychelles warblers (*Acrocephalus seychellensis*) [157], house sparrows (*Passer domesticus*)[158]). However, no mating preferences in relation to MHC variation was uncovered among great reed warblers (*Acrocephalus arundinaceus*)[159]. Both the savannah sparrow and the Seychelles warbler are highly promiscuous species[160,161], while the house sparrow and great reed warbler are not[162-164]. Patterns of mating preferences for MHC dissimilar mates were documented in the highly promiscuous savannah sparrow and Seychelles warbler, which is consistent with the results provided in Paper IV, that MHC diversity should be elevated in promiscuous passerines. These previous studies regarding MHC related mate choice in passerines were based on intra-specific comparisons only. In Paper IV we provide an inter-specific study encompassing eight related species from two different families of passerines. The association between promiscuity and MHC diversity uncovered in Paper IV indicates that extrapair mate choice

is important for the evolution of MHC polymorphism in passerines. In that respect, our multispecies approach has facilitated a broader perspective on the evolution of female promiscuity, a large theme in modern evolutionary biology.

Future prospects

During the work of this thesis, important results have been revealed for the understanding of replication slippage dynamics. Our screening for *HrU10* slippage mutations is to our knowledge the first to describe mutational patterns of the same microsatellite marker in a multispecies approach (Paper II). Nevertheless, the generality of our results needs to be addressed more comprehensively, using several markers widespread throughout the genome. Unfortunately, such genetic markers with detectable mutation rates applicable in multiple species are difficult to obtain.

Today, we are facing a new era of sequence technology[165-167]. In the next decade, 'next generation sequencing' will revolutionize the amount of available sequence information[168], especially in non-model organisms. Thus, potential new markers for microsatellite mutation screening may be developed, both with respect to mutagenesis of interrupting motifs and comparative mutation studies concerning multiple loci and species.

The extensive polymorphism and complex gene organization uncovered in MHCIIB genes in passerine species (Paper III and Paper IV) make standard Sanger-sequencing[169] approaches challenging for genotyping protocols in this group of birds. Consequently, next generation sequencing holds a great potential for comprehensive assessments of passerine MHC genes[97]. The first studies utilizing high throughput sequencing on MHC genes in natural populations of non-model organisms have been published[170,171], and the first article using this technique to obtain MHC sequence data in a passerine bird will soon be published (W. Babik, pers. com)[172]. Hence, our knowledge regarding MHC polymorphism, MHC organization and the relationship between MHC and other ecological factors in passerine birds, will probably be significantly increased in the years to come.

We identified a positive association between MHC diversity and promiscuity among eight species of passerines (Paper IV), and suggest that promiscuity may be a mechanism creating MHC diversity in passerines birds. A diversifying selection pressure at the MHC genes may maintain pathogen resistance in a continuously adapting pathogenic regime. However, we have no data regarding specific pathogen exposure and MHC diversity. Therefore, future studies should address the variation in pathogenic exposure in relation to promiscuity levels and MHC polymorphism in multispecies examinations.

Literature cited

1 Griffiths AJF, Gelbart WM, Lewontin RC & Miller JH. *Modern genetic analyses - Integrating genes and genomes*. 2002. (W. H. Freeman and Company).

2 Mathews CK, Van Holde KE & Ahern KG. in *Biochemistry*. 2000. Ch. 4, 84-118 (Addison-Wesley Publishing Company).

3 Pearson H. Genetics: What is a gene? 2006. *Nature* **441**, 398-401.

4 Mendel G. Versuche über Plflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brünn. 1865. *Abhandlungen*, 3-47.

5 Costa FF. Non-coding RNAs: Meet thy masters. 2010. *Bioessays* **32**, 599-608.

6 Hentze MW & Kühn LC. Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by irin, nitric oxide, and oxidative stress. 1996. *Proc Natl Acad Sci USA* **93**, 8175-8182.

7 Gilbert W. Why genes in pieces? 1978. *Nature* **271**, 501-501.

8 McClintock B. The origin and behavior of mutable loci in maize. 1950. *Proc Natl Acad Sci USA* **36**, 344-355.

9 Vanin EF. Processed pseudogenes: Characteristics and evolution. 1985. *Annu Rev Genet* **19**, 253-272.

10 Blackburn EH & Szostak JW. The molecular structure of centromeres and telomeres. 1984. *Ann Rev Biochem* **53**, 163-194.

11 Goldstein DB & Schlötterer C. *Microsatellites: Evolution and applications*. 1999. 1-352 (Oxford University Press).

12 Ohno S. in *Evolution of genetic systems. Brookhaven symposia in biology*. 1972. Vol. 23 (ed H.H. Smith) 366-370 (Gordon and Breach).

13 Biemont C & Vieira C. Genetics: Junk DNA as an evolutionary force. 2006. *Nature* **443**, 521-524.

14 Hancock JM. in *Microsatellites: Evolution and applications*. 1999. (eds D. B. Goldstein & C. Schlötterer) 1-9 (Oxford University Press).

15 International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. 2001. *Nature* **409**, 860-921.

16 Manto M-U. The wide spectrum of spinocerebellar ataxias (SCAs). 2005. *The Cerebellum* **4**, 2-6.

17 Kashi Y & Soller M. in *Microsatellites: Applications and Evolution*. 1999. (eds D. Goldstein & C. Schlötter) 10-23 (Oxford University Press).

18 Li YC, Korol AB, Fahima T, Beiles A & Nevo E. Microsatellites: Genomic distribution, putative functions and mutational mechanisms: a review. 2002. *Mol Ecol* **11**, 2453-2465.

19 Schlötter C & Wiehe T. in *Microsatellites: Evolution and application*. 1999. (eds D. B. Goldstein & C. Schlötter) 238-248 (Oxford University Press).

20 Ellegren H. Microsatellites: Simple sequences with complex evolution. 2004. *Nat Rev Genet* **5**, 435-445.

21 Levinson G & Gutman GA. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. 1987. *Mol Biol Evol* **4**, 2032-2021.

22 Kimura M & Ohta T. Stepwise mutation model and distribution of allelic frequencies in a finite population. 1978. *Proc Natl Acad Sci USA* **75**, 2868-2872.

23 Ellegren H. Microsatellite mutations in the germline: Implications for evolutionary inference. 2000. *Trends Genet* **16**, 551-558.

24 Beck NR, Double MC & Cockburn A. Microsatellite evolution at two hypervariable loci revealed by extensive avian pedigrees. 2003. *Mol Biol Evol* **20**, 54-61.

25 Boyer JC, Hawk JD, Stefanovic L & Farber RA. Sequence-dependent effect of interruptions on microsatellite mutation rate in mismatch repair-deficient human cells. 2008. *Mutat Res* **640**, 89-96.

26 Brinkmann B, Klintschar M, Neuhuber F, Huhne J & Rolf B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. 1998. *Am J Hum Genet* **62**, 1408-1415.

27 Brohede J, Møller AP & Ellegren H. Individual variation in microsatellite mutation rate in barn swallows. 2004. *Mutat Res* **545**, 73-80.

28 Brohede J, Primmer CR, Møller AP & Ellegren H. Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. 2002. *Nucleic Acids Res* **30**, 1997-2003.

29 Crozier RH, Kaufmann B, Carew ME & Crozier YC. Mutability of microsatellites developed for the ant *Camponotus consobrinus*. 1999. *Mol Ecol* **8**, 271-276.

30 Freund AM, Bichara M & Fuchs RP. Z-DNA-forming sequences are spontaneous deletion hot spots. 1989. *Proc Natl Acad Sci USA* **86**, 7465-7469.

31 Lai Y & Sun F. The relationship between microsatellite slippage mutation rate and the number of repeat units. 2003. *Mol Biol Evol* **20**, 2123-2131.

32 Paun O & Hörandl E. Evolution of hypervariable microsatellites in apomictic polyploid lineages of *Ranunculus carpaticola*: Directional bias at dinucleotide loci. 2006. *Genetics* **174**, 387-398.

33 Primmer CR, Saino N, Møller AP & Ellegren H. Directional evolution in germline microsatellite mutations. 1996. *Nat Genet* **13**, 391-393.

34 Wierdl M, Dominska M & Petes TD. Microsatellite instability in yeast: Dependence on the length of the microsatellite. 1997. *Genetics* **146**, 769-779.

35 Klein J. *Natural history of the major histocompatibility complex*. 1986. (John Wiley & sons).

36 Kaufmann SHE. CD8+ T lymphocytes in intracellular microbial infections. 1988. *Immunol Today* **9**, 168-174.

37 Rammensee H-G, Falk K & Rötzschke O. MHC molecules as peptide receptors. 1993. *Curr Opin Immunol* **5**, 35-44.

38 Dengjel J *et al.* From the Cover: Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. 2005. *Proc Natl Acad Sci USA* **102**, 7922-7927.

39 Villadangos JA. Presentation of antigens by MHC class II molecules: Getting the most out of them. 2001. *Mol Immunol* **38**, 329-346.

40 Björkman PJ *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. 1987. *Nature* **329**, 506-512.

41 Brown JH *et al.* Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. 1993. *Nature* **364**, 33-39.

42 Hughes AL & Yeager M. Natural selection at major histocompatibility complex loci of vertebrates. 1998. *Annu Rev Genet* **32**, 415-435.

43 Gaudieri S, Dawkins RL, Habara K, Kulski JK & Gojobori T. SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. 2000. *Genome Res* **10**, 1579-1586.

44 Geraghty DE, Daza R, Williams LM, Vu Q & Ishitani A. Genetics of the immune response: identifying immune variation within the MHC and throughout the genome. 2002. *Immunol Rev* **190**, 69-85.

45 Hughes AL *et al.* Natural selection at the class II major histocompatibility complex loci of mammals [and discussion]. 1994. *Phil Trans R Soc B* **346**, 359-367.

46 Hughes AL & Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. 1988. *Nature* **335**, 167-170.

47 Spurgin LG & Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. 2010. *P Roy Soc B-Biol Sci* **277**, 979-988.

48 Piertney SB & Oliver MK. The evolutionary ecology of the major histocompatibility complex. 2005. *Heredity* **96**, 7-21.

49 Penn DJ & Potts WK. The evolution of mating preferences and major histocompatibility complex genes. 1999. *Am Nat* **153**, 145-164.

50 Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. 2005. *Front Zool* **2**, 16.

51 Bernatchez L & Landry C. MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? 2003. *J. Evol. Biol.* **16**, 363-377.

52 Milinski M. The major histocompatibility complex, sexual selection, and mate choice. 2006. *Annu Rew Ecol Evol Syst* **37**, 159-186.

53 Griffith SC, Owens IPF & Thuman KA. Extra pair paternity in birds: A review of interspecific variation and adaptive function. 2002. *Mol Ecol* **11**, 2195-2212.

54 Albrecht T, Kreisinger J & Piálek J. The strength of direct selection against female promiscuity is associated with rates of extrapair fertilizations in socially monogamous songbirds. 2006. *Am Nat* **167**, 739-744.

55 Arnqvist G & Kirkpatrick M. The evolution of infidelity in socially monogamous passerines: The strength of direct and indirect selection on extrapair copulation behavior in females. 2005. *Am Nat* **165**, 26-37.

56 Westneat DF & Stewart IRK. Extra-pair paternity in birds: Causes, correlates, and conflict. 2003. *Annu Rew Ecol Evol Syst* **34**, 365-396.

57 Møller A & Cuervo J. Minisatellite mutation rates increase with extra-pair paternity among birds. 2009. *BMC Evol Biol* **9**, 100.

58 Petrie M & Roberts G. Sexual selection and the evolution of evolvability. 2006. *Heredity* **98**, 198-205.

59 Bartosch-Harlid A, Berlin S, Smith NG, Møller AP & Ellegren H. Life history and the male mutation bias. 2003. *Evolution* **57**, 2398-2406.

60 Kirkpatrick M & Ryan MJ. The evolution of mating preferences and the paradox of the lek. 1991. *Nature* **350**, 33-38.

61 Møller AP & Cuervo JJ. Sexual selection, germline mutation rate and sperm competition. 2003. *BMC Evol Biol* **3**, 6.

62 Fossøy F, Johnsen A & Lifjeld JT. Multiple genetic benefits of female promiscuity in a socially monogamous passerine. 2008. *Evolution* **62**, 145-156.

63 Garvin JC, Abroe B, Pedersen MC, Dunn PO & Whittingham LA. Immune response of nestling warblers varies with extra-pair paternity and temperature. 2006. *Mol Ecol* **15**, 3833-3840.

64 Johnsen A, Andersen V, Sunding C & Lifjeld JT. Female bluethroats enhance offspring immunocompetence through extra-pair copulations. 2000. *Nature* **406**, 296-299.

65 Westerdahl H. Passerine MHC: Genetic variation and disease resistance in the wild. 2007. *J Ornithol* **148**, 469-477.

66 Kleven O, Jacobsen F, Robertson RJ & Lifjeld JT. Extrapair mating between relatives in the barn swallow: a role for kin selection? 2005. *Biol Lett* **1**, 389-392.

67 Whittingham LA & Lifjeld JT. Extra-pair fertilizations increase the opportunity for sexual selection in the monogamous house martin (*Delichon urbica*). 1995. *J Avian Biol* **26**, 283-288.

68 Augustin J, Blomqvist D, Szép T, Szabó Z & Wagner R. No evidence of genetic benefits from extra-pair fertilisations in female sand martins (*Riparia riparia*). 2007. *J Ornithol* **148**, 189-198.

69 Delmore KE *et al.* Sex allocation and parental quality in tree swallows. 2008. *Behav Ecol* **19**, 1243-1249.

70 Bollmer J, Vargas F & Parker P. Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*). 2007. *Immunogenetics* **59**, 593-602.

71 Ekblom R *et al.* Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). 2007. *Mol Ecol* **16**, 1439-1451.

72 Hughes C, Miles S & Walbroehl J. Support for the minimal essential MHC hypothesis: a parrot with a single, highly polymorphic MHC class II B gene. 2008. *Immunogenetics* **60**, 219-231.

73 Kaufman J *et al.* The chicken B locus is a minimal essential major histocompatibility complex. 1999. *Nature* **401**, 923-925.

74 Strand T, Westerdahl H, Höglund J, Alatalo R V & Siitari H. The Mhc class II of the black grouse (*Tetrao tetrix*) consists of low numbers of B and Y genes with variable diversity and expression. 2007. *Immunogenetics* **59**, 725-734.

75 Kaufman J, Völk H & Wallny H-J. A "minimal essential Mhc" and an "unrecognized Mhc": Two extremes in selection for polymorphism. 1995. *Immunol Rev* **143**, 63-88.

76 Balakrishnan C *et al.* Gene duplication and fragmentation in the zebra finch major histocompatibility complex. 2010. *BMC Biology* **8**, 29.

77 Bollmer JL, Dunn PO, Whittingham LA & Wimpee C. Extensive MHC class II B gene duplication in a passerine, the common yellowthroat (*Geothlypis trichas*). 2010. *J Hered* **101**, 448-460.

78 Hill AVS *et al.* Common West African HLA antigens are associated with protection from severe malaria. 1991. *Nature* **352**, 595-600.

79 Klein J, Sato A, Nagl S & O'Huigin C. Molecular trans-species polymorphism. 1998. *Annu Rev Ecol Syst* **29**, 1-21.

80 Ober C *et al.* HLA and mate choice in humans. 1997. *Am J Hum Genet* **61**, 497-504.

81 Yamazaki K *et al.* Control of mating preferences in mice by genes in the major histocompatibility complex. 1976. *J Exp Med* **144**, 1324-1335.

82 Potts WK, Manning CJ & Wakeland EK. Mating patterns in seminatural populations of mice influenced by MHC genotype. 1991. *Nature* **352**, 619-621.

83 Yamazaki K *et al.* Familial imprinting determines H-2 selective mating preferences. 1988. *Science* **240**, 1331-1332.

84 Wedekind C, Seebeck T, Bettens F & Paepke AJ. MHC-dependent mate preferences in humans. 1995. *P Roy Soc B-Biol Sci* **260**, 245-249.

85 Havlicek J & Roberts SC. MHC-correlated mate choice in humans: A review. 2009. *Psychoneuroendocrinology* **34**, 497-512.

86 Chaix R, Cao C & Donnelly P. Is mate choice in humans MHC-dependent? 2008. *PLoS Genetics* **4**.

87 Hedrick PW & Black FL. HLA and mate selection: No evidence in South Amerindians. 1997. *Am J Hum Genet* **61**, 505-511.

88 Aeschlimann PB, Häberli MA, Reusch TBH, Boehm T & Milinski M. Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. 2003. *Behav Ecol Sociobiol* **54**, 119-126.

89 Reusch TBH, Haberli MA, Aeschlimann PB & Milinski M. Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. 2001. *Nature* **414**, 300-302.

90 Eizaguirre C, Yeates SE, Lenz TL, Kalbe M & Milinski M. MHC-based mate choice combines good genes and maintenance of MHC polymorphism. 2009. *Mol Ecol* **18**, 3316-3329.

91 Bos DH, Williams RN, Gopurenko D, Bulut Z & Dewoody JA. Condition-dependent mate choice and a reproductive disadvantage for MHC-divergent male tiger salamanders. 2009. *Mol Ecol* **18**, 3307-3315.

92 Olsson M *et al.* Major histocompatibility complex and mate choice in sand lizards. 2003. *Biol Lett* **270**, 254-256.

93 von Schantz T, Wittzell H, Goransson G, Grahn M & Persson K. MHC genotype and male ornamentation: Genetic evidence for the Hamilton-Zuk model. 1996. *P Roy Soc B-Biol Sci* **263**, 265-271.

94 Ekblom R, Sæther SA, Fiske P, Kålås J & Höglund J. Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe. 2010. *Genetica* **138**, 453-461.

95 Gillingham MAF *et al.* Cryptic preference for MHC-dissimilar females in male red junglefowl, Gallus gallus. 2009. *P Roy Soc B-Biol Sci* **276**, 1083-1092.

96 Saiki R *et al.* Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. 1985. *Science* **230**, 1350-1354.

97 Babik W. Methods for MHC genotyping in non-model vertebrates. 2010. *Mol Ecol Res* **10**, 237-251.

98 Meyerhans A, Vartanian J-P & Wain-Hobson S. DNA recombination during PCR. 1990. *Nucleic Acids Res* **18**, 1687-1691.

99 Bryja J, Galan M, Charbonnel N & Cosson JF. Analysis of major histocompatibility complex class II gene in water voles using capillary electrophoresis-single stranded conformation polymorphism. 2005. *Mol Ecol Notes* **5**, 173-176.

100 Lenz TL & Becker S. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - Implications for evolutionary analysis. 2008. *Gene* **427**, 117-123.

101 Judo MSB, Wedel AB & Wilson C. Stimulation and suppression of PCR-mediated recombination. 1998. *Nucleic Acids Res* **26**, 1819-1825.

102 Fischer SG & Lerman LS. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: Correspondance with melting theory. 1983. *Proc Natl Acad Sci USA* **80**, 1579-1583.

103 Orita M, Iwahana H, Kanazawa H, Hayashi K & Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphism. 1989. *Proc Natl Acad Sci USA* **86**, 2766-2770.

104 Arguello JR *et al.* Mutation detection and typing of polymorphic loci through double-strand conformation analysis. 1998. *Nat Genet* **18**, 192-194.

105 Canal D, Alcaide M, Anmarkrud J & Potti J. Towards the simplification of MHC typing protocols: Targeting classical MHC class II genes in a passerine, the pied flycatcher *Ficedula hypoleuca*. 2010. *BMC Res Notes* **3**, 236.

106 Freeman S & Herron JC. in *Evolutionary analysis*. 2004. (ed J. Challice) 109-140 (Pearson Education, Inc.).

107 Primmer CR, Møller AP & Ellegren H. New microsatellites from the pied flycatcher *Ficedula hypoleuca* and the swallow *Hirundo rustica* genomes. 1996. *Hereditas* **124**, 281-283.

108 Dawson DA *et al.* A predicted microsatellite map of the passerine genome based on chicken-passerine sequence similarity. 2006. *Mol Ecol* **15**, 1299-1320.

109 Riley DE, Jeon JS & Krieger JN. Simple repeat evolution includes dramatic primary sequence changes that conserve folding potential. 2007. *Biochem Biophys Res Commun* **355**, 619-625.

110 Crow JF & Kimura M. *An introduction to population genetics theory*. 1970. (Harper & Row).

111 Di Rienzo A *et al.* Mutational processes of simple-sequence repeat loci in human populations. 1994. *Proc Natl Acad Sci USA* **91**, 3166-3170.

112 Kimura M & Crow JF. The number of alleles that can be maintained in a finite population. 1964. *Genetics* **49**, 725-738.

113 Estoup A, Jarne P & Cornuet JM. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. 2002. *Mol Ecol* **11**, 1591-1604.

114 Burt A. Perspective: The evolution of fitness. 1995. *Evolution* **49**, 1-8.

115 Jeffreys AJ *et al.* Complex gene conversion events in germline mutation at human minisatellites. 1994. *Nat Genet* **6**, 136-145.

116 Primmer CR, Saino N, Møller AP & Ellegren H. Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows. 1998. *Mol Biol Evol* **15**, 1047-1054.

117 Bell SP & Dutta A. DNA replication in eukaryotic cells. 2002. *Ann Rev Biochem* **71**, 333-374.

118 Hunt PA & Hassold TJ. Sex matters in meiosis. 2002. *Science* **296**, 2181-2183.

119 Eward KL *et al.* DNA replication licensing in somatic and germ cells. 2004. *J Cell Sci* **117**, 5875-5886.

120 Shimmin LC, Chang BH-J & Li W-H. Male-driven evolution of DNA sequences. 1993. *Nature* **362**, 745-747.

121 Bohossian HB, Skaletsky H & Page DC. Unexpectedly similar rates of nucleotide substitution found in male and female hominids. 2000. *Nature* **406**, 622-625.

122 Ellegren H. Characteristics, causes and evolutionary consequences of male-biased mutation. 2007. *P Roy Soc B-Biol Sci* **274**, 1-10.

123 Hurst LD & Ellegren H. Sex biases in the mutation rate. 1998. *Trends Genet* **14**, 446-452.

124 Makova KD & Li W-H. Strong male-driven evolution of DNA sequences in humans and apes. 2002. *Nature* **416**, 624-626.

125 Ellegren H. Human mutation-blame (mostly) men. 2002. *Nat Genet* **31**, 9-10.

126 Hurst LD & Ellegren H. Human genetics: Mystery of the mutagenic male. 2002. *Nature* **420**, 365-366.

127 Tiemann-Boege I *et al.* The observed human sperm mutation frequency cannot explain the achondroplasia paternal age effect. 2002. *Proc Natl Acad Sci U S A* **99**, 14952-14957.

128 Miyata T, Hayashida H, Kuma K, Mitsuyasu K & Yasunaga T. Male-driven molecular evolution: A model and nucleotide sequence analysis. 1987. *Cold Spring Harb Symb Quant Biol* **52**, 863-867.

129 Haldane JBS. The mutation rate of the gene for haemophilia, and its segregation ratios in males and females. 1947. *Annals of Eugenics* **13**, 262-271.

130 Berlin S *et al.* Substitution rate heterogeneity and the male mutation bias. 2006. *J Mol Evol* **62**, 226-233.

131 Ellegren H & Fridolfsson A-K. Male-driven evolution of DNA sequences in birds. 1997. *Nat Genet* **17**, 182-184.

132 Ellegren H. Heterogeneous mutation processes in human microsatellite DNA sequences. 2000. *Nat Genet* **24**, 400-402.

133 Weber JL & Wong C. Mutation of human short tandem repeats. 1993. *Hum Mol Genet* **2**, 1123-1128.

134 Hoekert WE, Neufeglise H, Schouten AD & Menken SB. Multiple paternity and female-biased mutation at a microsatellite locus in the olive ridley sea turtle (*Lepidochelys olivacea*). 2002. *Heredity* **89**, 107-113.

135 Ortego J, Aparicio JM, Cordero PJ & Calabuig G. Characteristics of loci and individuals are associated with germline microsatellite mutation rates in lesser kestrels (*Falco naumanni*). 2008. *Mutat Res* **648**, 82-86.

136 Wittzell H, Bernot A, Auffray C & Zoorob R. Concerted evolution of two Mhc class II B loci in pheasants and domestic chickens. 1999. *Mol Biol Evol* **16**, 479-490.

137 Kikkawa E *et al.* Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus Spheniscus). 2009. *Immunogenetics* **61**, 341-352.

138 Alcaide M, Edwards SV, Negro JJ, Serrano D & Tella JL. Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). 2008. *Mol Ecol* **17**, 2652-2665.

139 Burri R, Niculita-Hirzel H, Roulin A & Fumagalli L. Isolation and characterization of major histocompatibility complex (MHC) class II B genes in the Barn owl (Aves: *Tyto alba*). 2008. *Immunogenetics* **60**, 543-550.

140 Alcaide M, Edwards SV & Negro JJ. Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. 2007. *J Mol Evol* **65**, 541-554.

141 Burri R, Hirzel HN, Salamin N, Roulin A & Fumagalli L. Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. 2008. *Mol Biol Evol* **25**, 1180-1191.

142 Bonneaud C *et al.* Diversity of Mhc class I and II B genes in house sparrows (*Passer domesticus*). 2004. *Immunogenetics* **55**, 855-865.

143 Freeman-Gallant CR, Johnson EM, Saponara F & Stanger M. Variation at the major histocompatibility complex in savannah sparrows. 2002. *Mol Ecol* **11**, 1125-1130.

144 Jarvi SI, Tarr CL, McIntosh CE, Atkinson CT & Fleischer RC. Natural selection of the major histocompatibility complex (Mhc) in Hawaiian honeycreepers (Drepanidinae). 2004. *Mol Ecol* **13**, 2157-2168.

145 Sato A *et al.* Evolution of Mhc class II B genes in Darwin's finches and their closest relatives: Birth of a new gene. 2001. *Immunogenetics* **53**, 792-801.

146 Vincek V *et al.* How large was the founding population of Darwin's finches? 1997. *P Roy Soc B-Biol Sci* **264**, 111-118.

147 Martinsohn JT, Sousa AB, Guethlein LA & Howard JC. The gene conversion hypothesis of MHC evolution: A review. 1999. *Immunogenetics* **50**, 168-200.

148 Richman AD, Herrera LG, Nash D & Schierup MH. Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*. 2003. *Genet Res Camb* **82**, 89-99.

149 Edwards SV & Dillon M. Hitchhiking and recombination in birds: Evidence from Mhc-linked and unlinked loci in red-winged blackbirds (*Agelaius phoeniceus*). 2005. *Genet Res Camb* **84**, 175-192.

150 Hunt HD, Pharr GT & Bacon LD. Molecular analysis reveals MHC class I intra-locus recombination in the chicken. 1994. *Immunogenetics* **40**, 370-375.

151 Miller HC & Lambert DM. Gene duplication and gene conversion in class II MHC genes of New Zealand robins (*Petroicidae*). 2004. *Immunogenetics* **56**, 178-191.

152 Silva M & Edwards S. Structure and evolution of a new avian MHC class II B gene in a sub-Antarctic seabird, the thin-billed prion (Procellariiformes: *Pachyptila belcheri*). 2009. *J Mol Evol*.

153 Warren WC *et al.* The genome of a songbird. 2010. *Nature* **464**, 757-762.

154 Nei M & Rooney AP. Concerted and birth-and-death evolution of multigene families. 2005. *Annu Rev Genet* **39**, 121-152.

155 Nei M, Gu X & Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. 1997. *Proc Natl Acad Sci USA* **94**, 7799-7806.

156 Freeman-Gallant CR, Meguerdichian M, Wheelwright NT & Sollecito SV. Social pairing and female mating fidelity predicted by restriction fragment length polymorphism similarity at the major histocompatibility complex in a songbird. 2003. *Mol Ecol* **12**, 3077-3083.

157 Richardson DS, Komdeur J, Burke T & von Schantz T. MHC-based patterns of social and extra-pair mate choice in the Seychelles warbler. 2005. *P Roy Soc B-Biol Sci* **272**, 759-767.

158 Bonneaud C, Chastel O, Federici P, Westerdahl H & Sorci G. Complex Mhc-based mate choice in a wild passerine. 2006. *P Roy Soc B-Biol Sci* **273**, 1111-1116.

159 Westerdahl H. No evidence of an MHC-based female mating preference in great reed warblers. 2004. *Mol Ecol* **13**, 2465-2470.

160 Richardson DS, Jury FL, Blaakmeer K, Komdeur J & Burke T. Parentage assignment and extra-group paternity in a cooperative breeder: the Seychelles warbler (*Acrocephalus sechellensis*). 2001. *Mol Ecol* **10**, 2263-2273.

161 Freeman-Gallant CR, Wheelwright NT, Meiklejohn KE & Sollecito SV. Genetic similarity, extrapair paternity, and offspring quality in Savannah sparrows (*Passerculus sandwichensis*). 2006. *Behav Ecol* **17**, 952-958.

162 Cordero PJ, Heeb P, Wetton JH & Parkin DT. Extra-pair fertilizations in tree sparrows *Passer montanus*. 2002. *Ibis* **144**, E67-E72.

163 Stewart IRK, Hanschu RD, Burke T & Westneat DF. Tests of ecological, phenotypic, and genetic correlates of extra-pair-paternity in the house sparrow. 2006. *The Condor* **108**, 399-413.

164 Hasselquist D, Bensch S & Von Schantz T. Low frequency of extrapair paternity in the polygynous great reed warbler, *Acrocephalus arundinaceus*. 1995. *Behav Ecol* **6**, 27-38.

165 Shendure J *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. 2005. *Science* **309**, 1728-1732.

166 Margulies M *et al.* Genome sequencing in microfabricated high-density picolitre reactors. 2005. *Nature* **437**, 376-380.

167 Metzker ML. Sequencing technologies - the next generation. 2010. *Nat Rev Genet* **11**, 31-46.

168 Schuster SC. Next-generation sequencing transforms today's biology. 2008. *Nat Meth* **5**, 16-18.

169 Sanger F, Nicklen S & Coulson AR. DNA sequencing with chain-terminating inhibitors. 1977. *Proc Natl Acad Sci USA* **74**, 5463-5467.

170 Galan M, Guivier E, Caraux G, Charbonnel N & Cosson J-F. A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. 2010. *BMC Genomics* **11**, 296.

171 Babik W, Taberlet P, Ejsmond MJ & Radwan J. New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. 2009. *Mol Ecol Res* **9**, 713-719.

172 Zagalska-Neubauer M *et al.* in *12th Congress of European Society for Evolutionary Biology* (Turin, 2009).

I

BMC Evolutionary Biology

Research article

Open Access

Microsatellite evolution: Mutations, sequence variation, and homoplasy in the hypervariable avian microsatellite locus *HrU10*

Jarl A Anmarkrud*, Oddmund Kleven, Lutz Bachmann and Jan T Lifjeld

Address: National Centre for Biosystematics, Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, NO-0318 Oslo, Norway

Email: Jarl A Anmarkrud* - j.a.anmarkrud@nhm.uio.no; Oddmund Kleven - oddmund.kleven@nhm.uio.no; Lutz Bachmann - lutz.bachmann@nhm.uio.no; Jan T Lifjeld - j.t.lifjeld@nhm.uio.no

* Corresponding author

Published: 9 May 2008

BMC Evolutionary Biology 2008, **8**:138 doi:10.1186/1471-2148-8-138

Received: 15 February 2008
Accepted: 9 May 2008

This article is available from: http://www.biomedcentral.com/1471-2148/8/138

Abstract

Background: Microsatellites are frequently used genetic markers in a wide range of applications, primarily due to their high length polymorphism levels that can easily be genotyped by fragment length analysis. However, the mode of microsatellite evolution is yet not fully understood, and the role of interrupting motifs for the stability of microsatellites remains to be explored in more detail. Here we present a sequence analysis of mutation events and a description of the structure of repeated regions in the hypervariable, pentanucleotide microsatellite locus *HrU10* in barn swallows (*Hirundo rustica*) and tree swallows (*Tachycineta bicolor*).

Results: In a large-scale parentage analysis in barn swallows and tree swallows, broods were screened for mutations at the *HrU10* locus. In 41 cases in the barn swallows and 15 cases in the tree swallows, mutations corresponding to the loss or gain of one or two repeat units were detected. The parent and mutant offspring alleles were sequenced for 33 of these instances (26 in barn swallows and 7 in tree swallows). Replication slippage was considered the most likely mutational process. We tested the hypothesis that *HrU10*, a microsatellite with a wide allele size range, has an increased probability of introductions of interruptive motifs (IMs) with increasing length of the repeated region. Indeed, the number and length of the IMs was strongly positively correlated with the total length of the microsatellite. However, there was no significant correlation with the length of the longest stretch of perfectly repeated units, indicating a threshold level for the maximum length of perfectly repeated pentanucleotide motifs in stable *HrU10* alleles. The combination of sequence and pedigree data revealed that 15 barn swallow mutations (58%) produced alleles that were size homoplasic to other alleles in the data set.

Conclusion: Our results give further insights into the mode of microsatellite evolution, and support the assumption of increased slippage rate with increased microsatellite length and a stabilizing effect of interrupting motifs for microsatellite regions consisting of perfect repeats. In addition, the observed extent of size homoplasy may impose a general caution against using hypervariable microsatellites in genetic diversity measures when alleles are identified by fragment length analysis only.

Background

Microsatellites consist of tandemly repeated sequence motifs, no more than 6 bases long. They are scattered throughout most eukaryotic genomes and are extensively used as tools for a wide range of applications, such as e.g. molecular forensics, parentage testing, analysis of genetic structure of populations and the assessment of phylogenetic relationships [1]. The major characteristic that makes microsatellites a useful and powerful genetic tool is the extensive length polymorphism that first of all reflects allelic variation in the number of the tandemly arranged perfect repeats [2]. However, "interrupting motifs" (IMs) that deviate in sequence from the repeated motif and mutations in the flanking regions may also contribute to the observed length polymorphism [3].

The molecular mechanisms for the development of microsatellite variation are not completely understood. In general, microsatellites have a high mutation rate (10^{-2}–10^{-6}) as compared to point mutations in coding genes [4]. It is accepted that the most common mutational mechanism affecting microsatellites is replication slippage, a process involving a gain or contraction of one or more repeat units [5,6]. Other microsatellite mutations might be caused by unequal crossing over, nucleotide substitutions, or duplication events [7]. Many factors might be important for the mutational processes in microsatellites, such as e.g. allele size, motif size, gender, and G/C content [8-15]. Mutation patterns may also depend on the genomic context such as the particular location on a chromosome and functional potential of the transcribed product [9,16-18], as well as the effectiveness of mismatch repair enzymes [19,20]. Moreover, mutation rates in microsatellites are also affected by stabilization patterns and potential secondary structures [13,21].

Various models have been put forward to explain and predict the mutation processes that affect microsatellite evolution (reviewed in [22]). The infinite allele model (IAM) [23] assumes that microsatellite mutations may create an infinite number of repeated units and allelic states not present in the population. Under the stepwise mutation model (SMM) [24] microsatellite mutations have the same probability of gaining or contracting one repeat unit. Thus, this model also takes into account back mutations. The generalized stepwise model (GSM) or two phase model (TPM) is an extension of the SMM and considers the probability for a microsatellite mutation to involve more than one unit [25]. According to the K-allele model (KAM) [26] there are K allelic states and equal probabilities to mutate towards any of the other (K-1) alleles.

Most studies concerning mutational processes in microsatellites have focused on size variation among the alleles (electromorphs), and not sequence variation. Genetic approaches lacking sequence data may hide essential information, e.g. substitutions, patterns of IMs and size homoplasy [3]. Homoplasic microsatellite alleles are alleles similar in state (length), but different in descent. Accordingly, one can divide microsatellite homoplasy into two types: (1) microsatellite alleles identical in length, but not in sequence (indistinguishable by fragment length analyses), and (2) alleles identical in both length and sequences, but with different evolutionary history (only detectable through mutations documented in known pedigrees). Some theoretical studies have tried to address the impact of homoplasy on genetic diversity analyses. Navascues and Emerson [27] showed that homoplasy affects various theoretical models differently. Simulations with high mutation rates ($\geq 10^{-4}$) for chloroplast microsatellites for example indicated an underestimation of homoplasy. In contrast, Estoup et al. [22] concluded that size homoplasy is not a substantial problem for population genetic studies, except for highly mutable microsatellites with strong allele size constraints in large populations.

Gaining empirical evidence of mutational processes affecting mutational diversity within natural populations is a demanding task. This is especially true for non-model organisms with comparatively few markers available and long generation time. One promising approach to overcome these difficulties is to genotype a large number of individuals in a population with known pedigree, using genetic markers with high mutation rates. One such marker is the microsatellite locus *HrU10* [28] in the European barn swallow (*Hirundo rustica*). This marker has been mapped to chromosome 18 in the chicken (*Gallus gallus*) genome [29], yet nothing is known about its functional potential. By means of fragment length analysis, Brohede et al. [30] estimated a mutation rate of 1.56% (i.e. 15 mutations in 960 meiotic events) for this locus. Currently there is only one *HrU10* sequence retrievable from GenBank (Accession nr: X97562). This sequence consists of a long tandemly repeated pentanucleotide (5'-TTCTC-3') stretch followed by an IM of two T's, two further repeat units and a variable tail of other arrangements of pyrimidines. Fragment length analysis in European barn swallows revealed that the majority of the *HrU10* alleles are approximately of similar length as the *HrU10* sequence in GenBank [30]. However, some alleles were up to three times as long. These longest *HrU10* alleles correspond to roughly 100 tandemly repeated pentanucleotide units, a remarkably large size for a microsatellite [31]. Wierdl et al. [32] suggested microsatellite stability to be related to the length of the stretch of tandem repeats, and postulated that large microsatellites have an increased probability to realign in a misaligned confirmation during replication resulting in a higher mutation rate. Since then,

increased instability of long microsatellites have been confirmed for several microsatellite loci, including *HrU10* [30]. One may assume that, in comparison to a microsatellite with a shorter repeat motif, a pentanucleotide repeat may establish a larger spatial conformation during loop formation of a slippage event. Somewhat misaligned nucleotides might occur and further increase the probability for mutations. Thereby also IMs may be introduced.

The fragment length analysis by Brohede et al. [30] indicated several size classes of the *HrU10* alleles in European barn swallow. Because of the high mutation rate, the *HrU10* microsatellite locus is well suited for testing the hypotheses of longer microsatellites being more unstable and more likely to gain IMs. Accordingly, one expects a positive correlation between the number of IMs and allele length and an upper size limit for the number of perfectly repeated motifs. In the present study, we sequenced a subset of *HrU10* alleles of different size classes from North American barn swallow and tree swallow (*Tachycineta bicolor*) that could be related to mutations in pedigree analyses [33]. This approach provides sequences from both the parent and the mutant offspring, and allows to investigate the formation of homoplasic alleles and to estimate the order of magnitude of size homoplasy for the *HrU10* microsatellite locus.

Methods

Samples and Genetic Analyses

This study was based on samples previously collected for the purpose of paternity testing in Canadian populations of barn swallows and tree swallows. Both species are socially monogamous passerine birds, but with high levels of extrapair paternity [33,34]. The barn swallow samples consisted of those already reported by Kleven et al. [33] and additional samples collected during 2003 and 2004. Barn swallows were genotyped with six to nine polymorphic microsatellite markers, including *HrU10*, and a detailed description of the markers, their variability and the paternity determination are presented elsewhere [33].

Tree swallows were genotyped with three polymorphic microsatellite markers, including *HrU10*. In cases where one of the three markers showed an allelic mismatch between offspring and one of the putative parents, an additional triplet of microsatellite markers were analyzed to distinguish mutation events from extrapair paternity. Details about the microsatellite markers, their polymorphism and the parentage determination of tree swallows are provided as additional file [see Additional file 1].

Mutations were detected by comparing the genotype of the offspring with that of its biological parents. We only included individuals for whom the genotypes of both biological parents were available. Furthermore, to avoid the problem of non-amplifying alleles, we only included mutations in parents that were heterozygous at the *HrU10* locus. We assumed the smallest mutational change in allele size to be most likely in cases where more than one parental allele could be the progenitor allele. To verify observed mutation events, we amplified microsatellite fragments twice for the parents and offspring involved in these cases.

Sequencing

The *HrU10* locus was amplified according to the protocol described in [33]. In addition, the reverse primer HrU10-EXT-R2 (5'-GCTGCTGTTCGAGGAAATAA-3') was designed to improve sequencing of tree swallow alleles. To reduce time consumption and lab costs we optimized a simple isolation strategy for *HrU10* alleles that did not require cloning. Alleles with size differences > 10 base pairs (bp) were separated on MetaPhor® agarose gels (Cambrex, East Rutherford, NJ) or on standard SeaKem® LE Agarose (Cambrex) if the alleles differed > 50 bp in size. The allele of interest was subsequently purified with the Nucleospin® Extract II gel extraction Kit (Macherey-Nagel, Düren, Germany), and sequenced directly in both directions using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's recommendations on an ABI 3100 Genetic Analyser (Applied Biosystems). As a further control, the sequence lengths were compared to previous fragment length analysis of the microsatellite [33].

Results

Microsatellite Structure Uncovered by Sequence Analyses

Pedigree analysis based on 2076 meiotic events in barn swallows and 496 meiotic events in tree swallows revealed 41 mutations for the microsatellite locus *HrU10* in barn swallows and 15 in tree swallows. According to fragment length analyses all mutations involved a gain or loss of five or ten bp, i.e. equivalent to one or two repeat units, which is consistent with the assumption of replication slippage. This represents a slippage rate of 1.97×10^{-2} in barn swallows and 3.02×10^{-2} in tree swallows. No other types of indels were observed. Sixty-six *HrU10* alleles involved in the observed germ line mutation were sequenced, that is, 33 sets of the offspring and the donor parent (GenBank accession numbers EU295565–EU295630). There were no site (four colonies) or year (two years) effects on mutation rates in barn swallows (both $P > 0.1$). The tree swallows were only sampled for one year at one location.

Woodruff et al. [35] approached the issue of "clustered mutations", which implies that related individuals may inherit identical genetic changes, contrasting an assumption that mutations are independent events. In this respect six of the parental individuals gave rise to two mutations

in the same family in the barn swallow population. In these cases, identical lengths where observed only twice. Only one of these two incidents of identical mutant lengths gave adequate sequences of both mutants. Nevertheless, in the particular case (Mut12 and Mut18 [see Additional file 2]) where sequences were obtained from both mutants, the outcome of the two mutations was dissimilar, verifying that the mutations represent independent events.

The nucleotide sequences confirmed that there was a gain or a contraction of one or two pentanucleotide units in all the 33 germ lines (26 in barn swallows and seven in tree swallows [see Additional file 2]). No nucleotide substitutions or other indels were detected. The sequence data revealed several different IMs for the *HrU10* alleles in both species, all of which consisted of distinct rearrangements of 1–30 pyrimidine bases (T and C). No purine bases were detected in the pyrimidine rich strand of the *HrU10* microsatellite. The subsequent statistical analyses were performed on the barn swallow parental alleles only, as the number of alleles that gave adequate sequences in tree swallows was considered too low (n = 7) for statistical testing.

First, we tested the hypothesis that longer microsatellites are more unstable and will consequently contain more IMs. If such a correlation occurred, then one would expect an upper size limit for the length of perfect repeats, and, accordingly there should be no correlation between the length of the longest stretch of perfect repeats of a given *HrU10* allele and the number of IMs. Correlation tests between total microsatellite length and (1) number of IMs (Spearman: $r_s = 0.55$, $P = 0.003$), and (2) total number of nucleotides contributing to the IMs ($r_s = 0.54$, $P = 0.004$) were significant. No significant correlation was found between the longest stretch of perfect tandemly repeated units and (1) number of IMs ($r_s = 0.09$, $P = 0.66$) or (2) total number of nucleotides in the IMs ($r_s = 0.14$, $P = 0.5$) (Figure 1).

We further determined roughly in which part of the *HrU10* allele a slippage event had taken place. The 16 shortest alleles showed one long (17–29 units) and one or two short (1–2 units) stretches of perfectly repeated units that are separated by IMs. The 10 longest alleles showed two long (14–55 units) and up to three short (0–5 units) stretches of perfectly repeated units. All replication induced mutations could be attributed to the long (≥14 units) stretches of perfect repeats. In 21 of 26 alleles (81%), the slippage mutation affected the longest stretch of perfectly repeated units. It is noteworthy that the other 5 mutations that did not occur in the longest stretch of

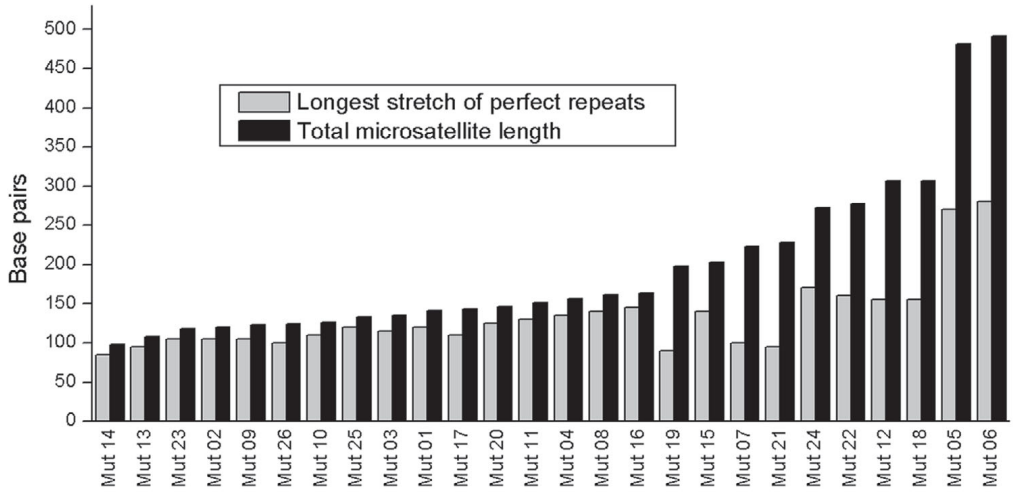

Figure 1
Correlation of array length of perfect repeats and *HrU10* microsatellite length. Plot of the longest stretch of perfectly repeated units and the corresponding total *HrU10* microsatellite length in sequenced mutant barn swallow alleles.

perfect repeats, all affected the alleles that were among the 10 longest.

In the barn swallow, 15 of the 26 (58%) observed mutations resulted in mutant alleles that were homoplasic to another sequenced allele in the data set. Nine (35%) of these mutations resulted in homoplasic alleles with respect to length, and six (23%), mutant alleles were also identical in sequence to another allele. No incidents of homoplasy due to mutations were detected in the tree swallows, but the number of sequences was significantly lower.

Features of the Mutations Revealed by the Fragment Length Analyses

The genotyped adult population consisted of 376 and 144 individuals in the barn and tree swallow populations, respectively. The allele size frequencies for the *HrU10* locus in the two species are illustrated in Figure 2 (note: 95 bp in the flanking regions of the microsatellite are not included in the presented sequences but in the fragment analyses). Due to seven non-amplifying alleles, only 745 barn swallow alleles were included. The median allele size of the *HrU10* microsatellite in the adult barn swallow population was 231 bp (± 2.0 SE, range = 175–581 bp, n = 704 alleles). Median size of the microsatellite of the mutant barn swallow alleles was 241 bp (± 13.8, range = 193–586, n = 41), i.e. significantly longer than for the entire population (Mann-Whitney U test: Z = -2.4, P = 0.02). The median allele size for the *HrU10* locus in the tree swallow population was 280 bp (± 4.6, range = 186–605, n = 272 alleles), and the median microsatellite length of the mutant tree swallow alleles was 323 bp (± 24.5, range = 222–518, n = 15). The mutation rates showed a tendency to be positively correlated with allele sizes for both barn swallows (GLM with binominal error distribution and logit link: χ^2_1 = 9.26, P = 0.002) and tree swallows (GLM with binominal error distribution and logit link: χ^2_1 = 3.35, P = 0.067). Estimations of mutation rates in relation to allele sizes are illustrated in Figure 3.

There were 26 (65%) expanding and 14 (35%) contracting mutations in the barn swallow population, which was not significantly different from equity (binominal test (two-tailed): P = 0.08), but may indicate a bias toward expansion. Directionality of the mutation was impossible to determine in one case because of a 5 bp difference to both possible parental alleles. No indication of directional bias for *HrU10* mutations was found in tree swallows with 7 (47%) expansions and 8 (53%) contractions (binominal test: P = 1.0). Directionality of mutations were not significantly affected by allele size neither in barn swallows (GLZ with binomial error distribution; χ^2 = 0.2, P = 0.7) nor in tree swallows (χ^2 = 0.1, P = 0.78). Furthermore, there was no positive correlation between length of the longest perfectly repeated core microsatellite unit and the directionality of the mutations in barn swallows (GLZ with binomial error distribution; χ^2 = 0.1, P = 0.81).

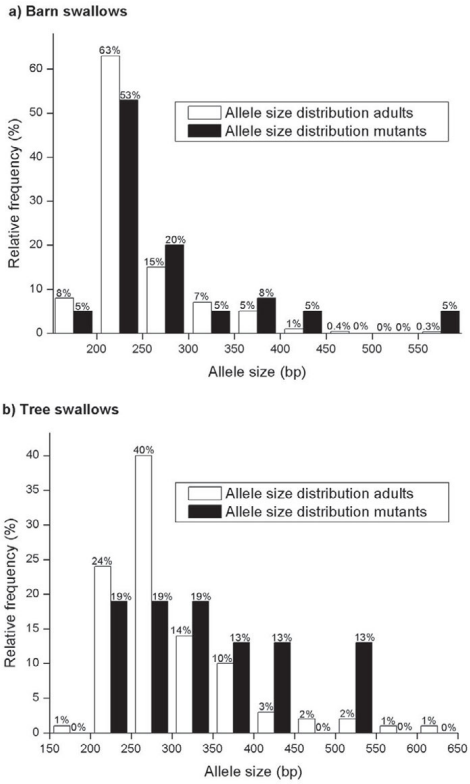

Figure 2
Size distribution of mutant and parental *HrU10* allele sizes. Size distribution of the *HrU10* locus in the adult population (white bars) and the mutant alleles (black bars) in **a)** barn swallows (n = 375) and **b)** tree swallows (n = 144). Each bar represents the alleles from the corresponding size class, which has been organized in groups of 50 and 50 bp.

In both species, there was a tendency for mutations to be maternally transmitted, as 66% (27/41) of the barn swallow mutations were observed in the female germ line (binominal test: P = 0.06) and 80% (12/15) in tree swallows (binominal test: P = 0.035).

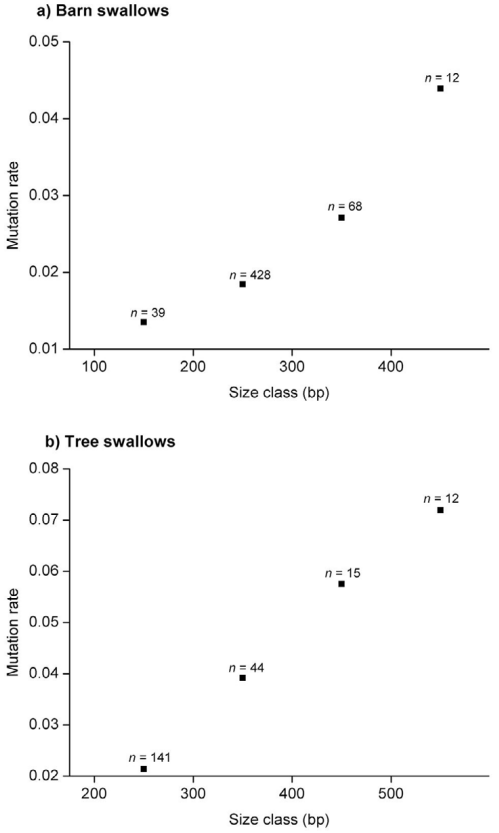

Figure 3
Correlation between mutation rate and *HrU10* allele sizes. Relationship between mutation rate and length of genotyped alleles in the population of biological parents in **a)** barn swallows and **b)** tree swallows. The alleles were lumped into 4 size classes (barn swallows: class 1 = 100–199 bp, class 2 = 200–299 bp, class 3 = 300–399 bp, class 4 = 400+ bp; tree swallows: class 1 = 200–299 bp, class 2 = 300–399 bp, class 3 = 400–499 bp, class 4 = 500+ bp). *n* refers to number of alleles in the particular size class.

Discussion

The sequence analyses presented in this study document two major features of the *HrU10* locus: (1) Frequent introductions of IMs which were strongly positively correlated with allele size and (2), frequent generation of homoplasic alleles (>50% of the mutations in the sequenced data set).

Evolution of HrU10 *– Introductions of IMs and Mutational Patterns*

All IMs consisted of different arrangements of Cs and Ts. No introduction of purines on the pyrimidine-rich strand was observed. The repeated motif of the *HrU10* locus is a pentanucleotide, and its spatial conformation during a slippage event may involve a larger loop compared to a smaller repeat unit. This may indicate that all the *HrU10* IMs are results of uncompleted replication slippage. If so, one needs to assume that the loop on the nascent or template strand will not include an entire unit before it realigns. This hypothesis is in line with all the IMs detected for *HrU10*, except for the $(TTCCC)_6$-repeat in Mut5 and Mut6 [see Additional file 2], which happened to be the longest *HrU10* alleles sequenced in the barn swallows. These two mutants have originated from the same parental allele that is 3.5 times longer than the median allele length.

An interesting feature related to the IMs, is the relatively constant maximal length of stretches of perfectly repeated core units, indicating a threshold length for stable arrays of perfectly repeated microsatellite units (Figure 1). One explanation for such a threshold level may be a selection pressure to retain stabile conformations. Such a selection pressure has been suggested to be important for preserving folding potential in repeated di-nucleotides [21,36] and tri-nucleotides [37]. In this respect, nothing is currently known for pentanucleotides.

Because the longest stretch of perfectly repeated units contributes most to the total microsatellite length in the *HrU10* alleles with only one long stretch of perfect tandem repeats, it is not surprising that the slippage events were restricted to this region of the microsatellite. Furthermore, theoretical and experimental approaches have suggested that slippage rates are approximately zero in very short repeated regions [38-41], supporting a low probability for slippages to occur in the short stretches in the *HrU10* alleles. Nevertheless, if length of perfectly repeated tandem units is more crucial for microsatellite instability than total microsatellite length, one would predict a consistent bias for slippage to affect the longest stretch of alleles comprising at least two long (>14 units) stretches of perfect core units. However, only 50% (5/10) of the mutations were introduced in the longer of the two stretches. This result is in agreement with a theory of total length of entire microsatellite being a more important factor for microsatellite stability than longest motif of perfect tandemly units. The observed threshold level for length of perfectly repeated units before introductions of IMs might then be explained by incomplete slippage (described above) being an increasingly more important mutational mechanism as the allele is destabilized due to growth. However, the statistical power for such a conclusion is rel-

atively poor. It is noteworthy that an opposite pattern was detected in an *in vitro* system of mono- and dinucleotides in a mutation study of human cell lineages [38]. Selection pressure for the maintenance of IMs has also been proposed important for tri-nucleotide repeat stability in genes related to various types of spinocerebellar ataxia [16,37,42].

The results from the fragment length analyses also support the prediction of long *HrU10* alleles to be more unstable, as the probability of a slippage event was positively related to allele sizes. The mutation rates were estimated to be more than three-fold higher in the longest compared with the smallest allele classes in the two swallow populations (Figure 3). Higher mutation rates for longer repeated regions have been reported in a number of previous studies [5,10,11,30,32,36,38,39,43-46]. The higher mutability in longer alleles can be explained by stabilization patterns concerning the mismatch-repair system, which may be less effective and as a result generate a relatively large probability for insertion of slippage events if the repeated region is sufficiently long, as suggested by Wierdl et al. [32]. The logic in this theory is that the number of possible conformations increases proportionally with the increase of repeated microsatellite units. However, if this hypothesis was correct, the microsatellite could, in theory, obtain uncontrolled growth and finally strike a large region on the chromosome. Nevertheless, microsatellites seems to have an upper size limit that rarely exceed 50 repeated units [31]. Still, some of the *HrU10* alleles uncovered in our study include almost 100 pentanucleotide units.

It has been suggested that short microsatellites tend to gain additional units whereas long microsatellites are more likely to lose units during a mutation event [39,47]. However, the support for this hypothesis is ambiguous. Primmer et al. [11], Eckert et al. [13] and Vigouroux et al. [48] have reported an overall directionality bias of slippage leading to expansion. Our data showed a similar tendency for barn swallow, though not statistically significant. Xu et al. [49] showed a constant slippage rate for expansion and an increasing slippage rate for contraction with increasing allele size, which is consistent with a general observation of an upper size limit for repeated regions. However, our data on the *HrU10* locus do not support that slippage directionality is length dependent, neither for total length of perfectly repeated core units nor for the total length of the entire microsatellite region. This result is in agreement with earlier studies for the *HrU10* locus [30].

Weber and Wong [50]; Garza et al. [51]; and Primmer et al. [52] have put forward hypotheses for the upper size limit for microsatellites. They postulate that the contraction of microsatellites is caused by large deletions that occur when a microsatellite reaches its maximal length potential. An example of such large deletion has been reported by Colson and Goldstein [53] who reported one incidence of an absent microsatellite in one allele of the *U1951* locus in *Drosophila melanogaster*. Other examples include a 27 units contraction of a dinucleotide microsatellite in *Ranunculus carpaticola* [46], and a 18 units contraction in a tetranucleotid repeat in superb fairy-wrens (*Malurus cyaneus*) [43]. It has also been suggested that if the balance between slippage and point mutations favours point mutations within the repeated region, the mutations may interrupt the feature of the microsatellite without enhancing large contractions [54], and eventually give rise to new diversity. Kruglyak et al. [55] developed a Markov chain model which confirmed that infinite microsatellite growth can be disabled by introductions of point mutations. In the sequence data presented here, all sequences contain IMs. These must have been introduced by other mutation mechanisms than regular slippage of entire units.

Because of the higher number of mitotic cell divisions in male than in female germ lines, it is plausible to expect that evolution of microsatellites, to some extent, is male-driven [14,15]. However, we observed a bias for mutations at the *HrU10* locus to be maternally transmitted in both barn swallows and tree swallows. Similar results have been reported by Brohede et al. [30] who observed a 2.5–5 fold increase in slippage rates in several hypermutable markers in females barn swallows compared to males, including the *HrU10* microsatellite. Beck et al. [43] also uncovered a bias favouring maternally transmitted slippages for one locus in the superb fairy-wren (*Malurus cyaneus*).

Cases of Size Homoplasy

More than 50% of the *HrU10* mutations sequenced in this study resulted in a mutant allele size homoplasic to another sequenced allele in the dataset, and 23% of the mutations resulted in alleles homoplasic in both allele length and sequence (identical alleles). Such homoplasy is only detectable through sequencing of observed mutations in known pedigrees and has to our knowledge not earlier been confirmed by empirical data. Size homoplasy may be problematic according to Estoup et al. [22] in instances with "(1) high mutation rates and (2) large population sizes together with (3) strong allele size constraints". The mutation rates observed in this study (1.97×10^{-2} per meiosis in barn swallow, 3.02×10^{-2} per meiosis in tree swallow) are among the highest ever reported for microsatellites, and our estimate is also concordant with that provided by Brohede et al. [30] for the same locus. A total number of 2070 meiotic events for the barn swallow population presented here is the largest data set ever

reported for an avian pentanucleotide microsatellite. Although it seems unlikely that there is a strong allele size constraint for the *HrU10* locus, our empirical results confirm that most of the observed mutations resulted in an electromorph already present among the 66 sequenced alleles. In consequence, homozygous individuals in terms of fragment length analyses are indeed not necessarily homozygous in terms of nucleotide sequences. Discrepancies between allelic variation detectable through fragment length analyses and sequence analyses have also been reported in other microsatellite studies [3,53,56]. Estoup et al. [22] approached the issue of size homoplasy theoretically based on frequently applied mutations models such as e.g. the SMM and the KAM. The proposed index of size homoplasy can be explained as the probability of two electromorphic alleles not being of common descent. However, these indexes have certain limitations when applied to the data on the *HrU10* locus provided here. First, the proposed homoplasy estimates relates to length and not sequence. Accordingly, there is no parameter that distinguishes between the two types of homoplasy. Second, the homoplasy index for the SMM does not include a parameter for the number of allelic states in a population, which is certainly crucial for homoplasy estimates. Third, the KAM accounts for allelic states, but assumes equal probability to mutate towards any of the other K-1 alleles. This is certainly not the case for the *HrU10* locus in barn swallows.

Many studies have focused on the occurrence of microsatellite size homoplasy within different taxa (e.g: humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*) [57], mammalian carnivores [58,59], birds [60,61], salmonids [3], pipefish (*Syngnathus typhle*) [62], crabs (*Limulus polyphemus*) [63], two bee species and the fresh water snail *Bulinus truncates* [64] and fruit flies (*Drosophila*) [65]). These examples support the notion that caution must be taken when microsatellite data are collected by electromorphic genotyping only. The high rate of mutations leading to size homoplasy in the present study provides support for alleles of identical size being attributed to common descent and hence causing bias in population genetic estimates.

Estoup et al[66] and Estoup and Cournet [67] suggested that the amount of size homoplasy is more important in interspecific than in intraspecific comparisons. Nevertheless, our results for *HrU10* provide evidence that homoplasy may play an important role also within populations. This conclusion is in agreement with the results published by van Oppen et al. [68], who found equally high amounts of homoplasy when comparing individuals among closely related taxa and among more distantly related species. *HrU10* is a frequently used marker for different genetic analyses of bird populations, especially parentage studies [33,69-72] because of the high allelic diversity enabling a powerful marker for parentage testing. However, our study indicates that *HrU10* should be used with caution whenever homoplasy may cause biased estimates of relatedness and genetic diversity.

Conclusion

Sequencing of 33 mutated and 66 parental *HrU10* alleles was consistent with the hypotheses that longer alleles tend to be more instable due to increased slippage rate. The observed positive correlation between the number of IMs and allele size supported the assumption of a threshold level for the maximal length of stable perfect repeats. Nonetheless, the particular location of the slippage positions in the mutated microsatellite alleles indicated that total microsatellite length is more important for microsatellite stability than the length of the longest stretch of perfect repeats. Mainly because of the high slippage rate, there is also a high level of homoplasy at the *HrU10* locus, i.e. 58% of the characterized mutations yielded an electromorph already present in the sequenced data set, including both type 1 and type 2 size homoplasy. The problem of size homoplasy imposes a general caution of using such hypermutable markers in fragment analyses assuming unique alleles by size only.

Authors' contributions

JAA carried out the sequence analyses, interpreted results and drafted the manuscript. OK performed the field work, carried out the fragment analyses, helped designing the study and drafting the manuscript. LB helped designing the study, interpreting results and drafting the manuscript. JTL initiated the study and helped drafting the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Analysis of paternity in tree swallows. A detailed description of the markers used for paternity analysis in the tree swallows.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-8-138-S1.pdf]

Additional file 2

Table presenting the sequenced HrU10 *microsatellites.* HrU10 *microsatellite sequences detected in* ***a)*** *barn swallows and* ***b)*** *tree swallows. The underlined nucleotides represent the core microsatellite unit (nucleotides not underlined are inserts of IMs). The regions in which a mutation occurred are depicted by red numbers. "bp" refers to total numbers of base pairs in the microsatellite region and "Type" refers to numbers of gains or contractions of units each mutation caused. All mutations which resulted in electromorphic identity with another sequenced allele are listed below the table.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-8-138-S2.pdf]

Acknowledgements

We are most grateful to professor Raleigh J. Robertson for all help and support for making our studies on barn swallows and tree swallows possible. We also thank all those that have contributed to the field and lab work. The study was funded by the Research Council of Norway (grant JTL) and the Natural History Museum, University of Oslo (PhD fellowship to JAA).

References

1. Goldstein DB, Schlötterer C: **Microsatellites: Evolution and applications.** Oxford, Oxford University Press; 1999. Page 352
2. Tautz D: **Hypervariabflity of simple sequences as a general source for polymorphic DNA markers.** *Nucl Acids Res* 1989, **17(16):**6463-6471.
3. Angers B, Bernatchez L: **Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information.** *Mol Biol Evol* 1997, **14(3):**230-238.
4. Li YC, Korol AB, Fahima T, Beiles A, Nevo E: **Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review.** *Mol Ecol* 2002, **11(12):**2453-2465.
5. Levinson G, Gutman GA: **Slipped-strand mispairing: a major mechanism for DNA sequence evolution.** *Mol Biol Evol* 1987, **4(3):**203-21.
6. Schlötterer C, Tautz D: **Slippage synthesis of simple sequence DNA.** *Nucleic Acids Res* 1992, **20(2):**211-215.
7. Hancock JM: **Microsatellites and other simple sequences: genomic context and mutational mechanisms.** In *Microsatellites: Evolution and applications* Edited by: Goldstein D, Schlötterer C. New York , Oxford University Press; 1999:1-9.
8. Anderson TJ, Su XZ, Roddam A, Day KP: **Complex mutations in a high proportion of microsatellite loci from the protozoan parasite Plasmodium falciparum.** *Mol Ecol* 2000, **9(10):**1599-1608.
9. Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R: **Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci.** *Proc Natl Acad Sci USA* 1997, **94(3):**1041-1046.
10. Brohede J, Møller AP, Ellegren H: **Individual variation in microsatellite mutation rate in barn swallows.** *Mutat Res* 2004, **545(1-2):**73-80.
11. Primmer CR, Saino N, Møller AP, Ellegren H: **Directional evolution in germline microsatellite mutations.** *Nat Genet* 1996, **13(4):**391-393.
12. Whittaker JC, Harbord RM, Boxall N, Mackay I, Dawson G, Sibly RM: **Likelihood-based estimation of microsatellite mutation rates.** *Genetics* 2003, **164(2):**781-787.
13. Eckert KA, Yan G, Hile SE: **Mutation rate and specificity analysis of tetranucleotide microsatellite DNA alleles in somatic human cells.** *Mol Carcinog* 2002, **34(3):**140-150.
14. Ellegren H: **Heterogeneous mutation processes in human microsatellite DNA sequences.** *Nat Genet* 2000, **24(4):**400-402.
15. Ellegren H: **Human mutation-blame (mostly) men.** *Nat Genet* 2002, **31(1):**9-10.
16. Tomiuk J, Bachmann L, Bauer C, Rolfs A, Schols L, Roos C, Zischler H, Schuler MM, Bruntner S, Riess O, Bauer P: **Repeat expansion in spinocerebellar ataxia type 17 alleles of the TATA-box binding protein gene: an evolutionary approach.** *Eur J Hum Genet* 2006, **15(1):**81-87.
17. Schlötterer C, Ritter R, Harr B, Brem G: **High mutation rate of a long microsatellite allele in Drosophila melanogaster provides evidence for allele-specific mutation rates.** *Mol Biol Evol* 1998, **15(10):**1269-1274.
18. Hawk JD, Stefanovic L, Boyer JC, Petes TD, Farber RA: **Variation in efficiency of DNA mismatch repair at different sites in the yeast genome.** *Proc Natl Acad Sci USA* 2005, **102(24):**8639-8643.
19. Modrich P, Lahue R: **Mismatch repair in replication fidelity, genetic recombination, and cancer biology.** *Annu Rev Biochem* 1996, **65:**101-133.
20. Harr B, Todorova J, Schlötterer C: **Mismatch repair-driven mutational bias in D. melanogaster.** *Mol Cell* 2002, **10(1):**199-205.
21. Riley DE, Jeon JS, Krieger JN: **Simple repeat evolution includes dramatic primary sequence changes that conserve folding potential.** *Biochem Biophys Res Commun* 2007, **355(3):**619-625.
22. Estoup A, Jarne P, Cornuet JM: **Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis.** *Mol Ecol* 2002, **11(9):**1591-1604.
23. Kimura M, Crow JF: **The number of alleles that can be maintained in a finite population.** *Genetics* 1964, **49:**725-738.
24. Kimura M, Ohta T: **Stepwise mutation model and distribution of allelic frequencies in a finite population.** *Proc Natl Acad Sci USA* 1978, **75(6):**2868-2872.
25. Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB: **Mutational processes of simple-sequence repeat loci in human populations.** *Proc Natl Acad Sci USA* 1994, **91(8):**3166-3170.
26. Crow JF, Kimura M: **An introduction to population genetics theory.** New York, Harper & Row; 1970:591.
27. Navascues M, Emerson BC: **Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy.** *Mol Ecol* 2005, **14(5):**1333-1341.
28. Primmer CR, Møller AP, Ellegren H: **New microsatellites from the pied flycatcher Ficedula hypoleuca and the swallow Hirundo rustica genomes.** *Hereditas* 1996, **124(3):**281-283.
29. Dawson DA, Burke T, Hansson B, Pandhal J, Hale MC, Hinten GN, Slate J: **A predicted microsatellite map of the passerine genome based on chicken-passerine sequence similarity.** *Mol Ecol* 2006, **15(5):**1299-1320.
30. Brohede J, Primmer CR, Møller AP, Ellegren H: **Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci.** *Nucleic Acids Res* 2002, **30(9):**1997-2003.
31. Buschiazzo E, Gemmell NJ: **The rise, fall and renaissance of microsatellites in eukaryotic genomes.** *Bioessays* 2006, **28(10):**1040-1050.
32. Wierdl M, Dominska M, Petes TD: **Microsatellite instability in yeast: dependence on the length of the microsatellite.** *Genetics* 1997, **146(3):**769-779.
33. Kleven O, Jacobsen F, Robertson RJ, Lifjeld JT: **Extrapair mating between relatives in barn swallow: a role for kin selection?** *Biol Lett* 2005, **1:**389-392.
34. Stapleton MK, Kleven O, Lifjeld JT, Robertson RJ: **Female tree swallows (Tachycineta bicolor) increase offspring heterozygosity through extrapair mating.** *Behav Ecol Sociobiol* 2007, **61(11):**1725-1733.
35. Woodruff RC, Hual H, Thompson JN: **Clusters of identical new mutation in the evolutionary landscape.** *Genetica* 1996, **98(2):**149-160.
36. Freund AM, Bichara M, Fuchs RP: **Z-DNA-forming sequences are spontaneous deletion hot spots.** *Proc Natl Acad Sci USA* 1989, **86(19):**7465-7469.
37. Pearson CE, Eichler EE, Lorenzetti D, Kramer SF, Zoghbi HY, Nelson DL, Sinden RR: **Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation.** *Biochemistry* 1998, **37(8):**2701-2708.
38. Boyer JC, Hawk JD, Stefanovic L, Farber RA: **Sequence-dependent effect of interruptions on microsatellite mutation rate in mismatch repair-deficient human cells.** *Mutat Res* in press.
39. Lai Y, Sun F: **The relationship between microsatellite slippage mutation rate and the number of repeat units.** *Mol Biol Evol* 2003, **20(12):**2123-2131.

40. Rose O, Falush D: **A threshold size for microsatellite expansion.** *Mol Biol Evol* 1998, **15(5):**613-615.
41. Sibly RM, Whittaker JC, Talbot M: **A Maximum-Likelihood Approach to Fitting Equilibrium Models of Microsatellite Evolution.** *Mol Biol Evol* 2001, **18(3):**413-417.
42. Sobczak K, Krzyzosiak WJ: **Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability.** *Human Mutation* 2004, **24(3):**236-247.
43. Beck NR, Double MC, Cockburn A: **Microsatellite evolution at two hypervariable loci revealed by extensive avian pedigrees.** *Mol Biol Evol* 2003, **20(1):**54-61.
44. Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B: **Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat.** *Am J Hum Genet* 1998, **62(6):**1408-1415.
45. Crozier RH, Kaufmann B, Carew ME, Crozier YC: **Mutability of microsatellites developed for the ant Camponotus consobrinus.** *Mol Ecol* 1999, **8(2):**271-276.
46. Paun O, Hörandl E: **Evolution of hypervariable microsatellites in apomictic polyploid lineages of Ranunculus carpaticola: directional bias at dinucleotide loci.** *Genetics* 2006, **174(1):**387-398.
47. Huang QY, Xu FH, Shen H, Deng HY, Liu YJ, Liu YZ, Li JL, Recker RR, Deng HW: **Mutation patterns at dinucleotide microsatellite loci in humans.** *Am J Hum Genet* 2002, **70(3):**625-634.
48. Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J: **Rate and pattern of mutation at microsatellite loci in maize.** *Mol Biol Evol* 2002, **19(8):**1251-1260.
49. Xu X, Peng M, Fang Z: **The direction of microsatellite mutations is dependent upon allele length.** *Nat Genet* 2000, **24(4):**396-399.
50. Weber JL, Wong C: **Mutation of human short tandem repeats.** *Hum Mol Genet* 1993, **2(8):**1123-1128.
51. Garza JC, Slatkin M, Freimer NB: **Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size.** *Mol Biol Evol* 1995, **12(4):**594-603.
52. Primmer CR, Saino N, Møller AP, Ellegren H: **Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows.** *Mol Biol Evol* 1998, **15:**1047-1054 [http://mbe.oxfordjournals.org/cgi/reprint/15/8/1047].
53. Colson I, Goldstein DB: **Evidence for complex mutations at microsatellite loci in Drosophila.** *Genetics* 1999, **152(2):**617-627.
54. Schug MD, Wetterstrand KA, Gaudette MS, Lim RH, Hutter CM, Aquadro CF: **The distribution and frequency of microsatellite loci in Drosophila melanogaster.** *Mol Ecol* 1998, **7(1):**57-70.
55. Kruglyak S, Durrett RT, Schug MD, Aquadro CF: **Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations.** *Proc Natl Acad Sci USA* 1998, **95(18):**10774-10778.
56. Bachmann L, Bareiss P, Tomiuk J: **Allelic variation, fragment length analyses and population genetic models: a case study on Drosophila microsatellites.** *J Zool Syst Evol Research* 2004, **42(3):**215-223.
57. Garza JC, Freimer NB: **Homoplasy for size at microsatellite loci in humans and chimpanzees.** *Genome Res* 1996, **6(3):**211-217.
58. Domingo-Roura X, Lopez-Giraldez F, Saeki M, Marmi J: **Phylogenetic inference and comparative evolution of a complex microsatellite and its flanking regions in carnivores.** *Genet Res* 2005, **85(3):**223-233.
59. Culver M, Menotti-Raymond MA, O'Brien SJ: **Patterns of size homoplasy at 10 microsatellite loci in pumas (Puma concolor).** *Mol Biol Evol* 2001, **18(6):**1151-1156.
60. Primmer CR, Ellegren H: **Patterns of molecular evolution in avian microsatellites.** *Mol Biol Evol* 1998, **15(8):**997-1008.
61. Shepherd LD, Lambert DM: **Mutational bias in penguin microsatellite DNA.** *J Hered* 2005, **96(5):**566-571.
62. Jones AG, Rosenqvist G, Berglund A, Avise JC: **Clustered microsatellite mutations in the pipefish Syngnathus typhle.** *Genetics* 1999, **152(3):**1057-1063.
63. Orti G, Pearse DE, Avise JC: **Phylogenetic assessment of length variation at a microsatellite locus.** *Proc Natl Acad Sci USA* 1997, **94(20):**10745-10749.
64. Viard F, Franck P, Dubois MP, Estoup A, Jarne P: **Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species.** *J Mol Evol* 1998, **47(1):**42-51.
65. Krafsur ES, Cummings MA, Endsley MA, Marquez JG, Nason JD: **Geographic differentiation in the house fly estimated by microsatellite and mitochondrial variation.** *J Hered* 2005, **96(5):**502-512.
66. Estoup A, Tailliez C, Cornuet JM, Solignac M: **Size homoplasy and mutational processes of interrupted microsatellites in two bee species, Apis mellifera and Bombus terrestris (Apidae).** *Mol Biol Evol* 1995, **12(6):**1074-1084.
67. Estoup A, Cornuet JM: **Microsatellite evolution: inferences from population data.** In *Microsatellites: evolution and applications* Edited by: Goldstein D, Schlötterer C. New York, Oxford University Press; 1999:49-65.
68. van Oppen MJH, Rico C, Turner GF, Hewitt GM: **Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids.** *Mol Biol Evol* 2000, **17(4):**489-498.
69. Møller AP, Brohede J, Cuervo JJ, de Lope F, Primmer C: **Extrapair paternity in relation to sexual ornamentation, arrival date, and condition in a migratory bird.** *Behav Ecol* 2003, **14(5):**707-712.
70. Neuman CR, Safran RJ, Lovette IJ: **Male tail streamer length does not predict apparent or genetic reproductive success in North American barn swallows Hirundo rustica erythrogaster.** *J Avian Biol* 2007, **38(1):**28-36.
71. O'Brien EL, Dawson RD: **Context-dependent genetic benefits of extra-pair mate choice in a socially monogamous passerine.** *Behav Ecol Sociobiol* 2007, **61:**775-782.
72. Whittingham LA, Dunn PO: **Survival of extrapair and within-pair young in tree swallows.** *Behav Ecol* 2001, **12(4):**496-500.

ELECTRONIC APPENDIX

This is the Electronic Appendix to the article

Microsatellite evolution: Mutations, sequence variation, and homoplasy in the hypervariable avian microsatellite locus *HrU10*

by

Jarl A. Anmarkrud, Oddmund Kleven, Lutz Bachmann, Jan T. Lifjeld

BMC Evolutionay Biology

Analysis of paternity in tree swallows

DNA was extracted from blood or tissue samples with a commercial kit (E.Z.N.A, Omega Bio-Tek, U.S.A.) and paternity analysed based on three highly polymorphic microsatellite markers through polymerase chain reaction (PCR). Each 10 µL reaction consisted of about 30 ng of genomic DNA, 0.5 µM of each primer (forward primers were fluorescently dyed), 0.1 mM dNTP mix (ABgene, U.K.) and 0.2 units of DNA polymerase (DyNAzyme, Finnzymes, Finland) in the manufacturer's buffer (final concentrations of 10 mM Tris HCl, 1.5 mM MgCl2, 50 mM KCl, 0.1% Triton X-100). PCR was run on a GeneAmp 9700 Thermocycler (Applied Biosystems, U.S.A.). The PCR profile used consisted of an initial denaturing step at 94°C for 5 min, followed by 35 cycles consisting of 94°C for 30 s, X°C (50°C for locus *Aar4*, and 55°C for loci *HrU6* and *HrU10*) for X s (30 s for locus *Aar4*, and 40 s for loci *HrU6* and *HrU10*), and 72°C for 30 s. The PCR profile was terminated with 72°C for 7 min followed by 4°C for 5 min. PCR products were sized using a capillary automated ABI 3100 sequencer (Applied Biosystems, U.S.A.) and the data was analyzed with GeneMapper v3.0 analytical software (Applied Biosystems, U.S.A.). Polymorphism of the microsatellite markers was calculated using CERVUS v3.0 [1] (see table 1). The combined exclusion probability for the three markers was higher than 0.99. Null-alleles were detected at one locus (*HrU10*) and mutations at two loci (*HrU6* and *HrU10*). Hence, to resolve cases of paternity uncertainty due to a single allelic mismatch between the genotype of a young and the putative parent, we used an additional triplet of microsatellite markers (*Ltr6*, *Pdoµ5* and *Tbi81*). Detailed information about the PCR protocol and polymorphism of these latter markers is presented elsewhere [2].

Table 1. Polymorphism among three microsatellite loci used for analysis of paternity in tree swallows (*Tachycineta bicolor*). Data on the number of alleles (*k*), number of genotyped adult individuals (*n*), observed heterozygosity (H_O), expected heterozygosity (H_E), probability of exclusion assuming no parents known (P_{ei}), probability of exclusion assuming one parent known (P_{ej}), and estimated frequency of null-alleles (N_e) are presented.

locus (references)	*k*	*n*	H_O	H_E	P_{ei}	P_{ej}	N_e
Aar4 [3]	15	146	0.88	0.89	0.62	0.77	+0.002
HrU6 [4]	81	146	0.92	0.92	0.71	0.83	−0.003
HrU10 [5]	142	144	0.84	0.99	0.96	0.98	+0.082

References

1. Kalinowski ST, Taper ML, Marshall TC: **Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment**. *Mol Ecol* 2007, **16**(5):1099-1106.
2. Stapleton MK, Kleven O, Lifjeld JT, Robertson RJ: **Female tree swallows (*Tachycineta bicolor*) increase offspring heterozygosity through extrapair mating**. *Behav Ecol Sociobiol* 2007, **161**:1725-1733.
3. Hansson B, Bensch S, Hasselquist D, Lillandt BG, Wennerberg L, von Schantz T: **Increase of genetic variation over time in a recently founded population of great reed warblers (*Acrocephalus arundinaceus*) revealed by microsatellites and DNA fingerprinting**. *Mol Ecol* 2000, **9**(10):1529-1538.
4. Primmer CR, Møller AP, Ellegren H: **Resolving genetic relationships with microsatellite markers: a parentage testing system for the swallow *Hirundo rustica***. *Mol Ecol* 1995, **4**(4):493-498.
5. Primmer CR, Møller AP, Ellegren H: **New microsatellites from the pied flycatcher *Ficedula hypoleuca* and the swallow *Hirundo rustica* genomes**. *Hereditas* 1996, **124**:281-283.

a) Barn swallow

Mut		Sequence	bp	Type
1	Parent	TCCTC(TTCTC)$_{24}$TT(TTCTC)$_2$(TC)$_2$	141	+1
	Offspring	TCCTC(TTCTC)$_{25}$TT(TTCTC)$_2$(TC)$_2$	146[a]	
2	Parent	TC(TTCTC)$_{21}$TTC(TTCTC)$_2$	120	+1
	Offspring	TC(TTCTC)$_{22}$TTC(TTCTC)$_2$	125	
3	Parent	(TC)$_2$(TTCTC)$_{23}$TT(TTCTC)$_2$(TC)$_2$	135	-1
	Offspring	(TC)$_2$(TTCTC)$_{22}$TT(TTCTC)$_2$(TC)$_2$	130	
4	Parent	TCCTC(TTCTC)$_{27}$TT(TTCTC)$_2$(TC)$_2$	156	-2
	Offspring	TCCTC(TTCTC)$_{25}$TT(TTCTC)$_2$(TC)$_2$	146[a]	
5	Parent	TCCTC(TTCTC)$_{32}$(TTCCC)$_6$(TTCTC)$_{55}$TT(TTCTC)$_2$(TC)$_2$	486[b]	+1
	Offspring	TCCTC(TTCTC)$_{32}$(TTCCC)$_6$(TTCTC)$_{56}$TT(TTCTC)$_2$(TC)$_2$	491[b]	
6	Parent	TCCTC(TTCTC)$_{32}$(TTCCC)$_6$(TTCTC)$_{55}$TT(TTCTC)$_2$(TC)$_2$	486[b]	-1
	Offspring	TCCTC(TTCTC)$_{32}$(TTCCC)$_6$(TTCTC)$_{54}$TT(TTCTC)$_2$(TC)$_2$	481[b]	
7	Parent	(TTCTC)$_4$TCCTC(TTCTC)$_{20}$TC(TTCTC)$_{17}$TT(TTCTC)(TC)$_2$	223[c]	+1
	Offspring	(TTCTC)$_4$TCCTC(TTCTC)$_{20}$TC(TTCTC)$_{18}$TT(TTCTC)(TC)$_2$	228[d]	
8	Parent	TCCTC(TTCTC)$_{28}$TT(TTCTC)$_2$(TC)$_2$	161	+1
	Offspring	TCCTC(TTCTC)$_{29}$TT(TTCTC)$_2$(TC)$_2$	166	
9	Parent	TC(TTCTC)$_{21}$TT(TTCTC)$_2$(TC)$_2$	123[e]	-1
	Offspring	TC(TTCTC)$_{20}$TT(TTCTC)$_2$(TC)$_2$	118[f]	
10	Parent	(TTCTC)$_{22}$TT(TTCTC)$_2$(TC)$_2$	126	-1
	Offspring	(TTCTC)$_{21}$TT(TTCTC)$_2$(TC)$_2$	121	
11	Parent	TCCTC(TTCTC)$_{26}$TT(TTCTC)$_2$(TC)$_2$	151[g]	-1
	Offspring	TCCTC(TTCTC)$_{25}$TT(TTCTC)$_2$(TC)$_2$	146[a]	
12	Parent	(TTCTC)$_{31}$(TTCCTC)(TTCTC)$_{26}$TT(TTCTC)$_2$(TC)$_2$	307[h]	+1
	Offspring	(TTCTC)$_{31}$(TTCCTC)(TTCTC)$_{27}$TT(TTCTC)$_2$(TC)$_2$	312[i]	
13	Parent	(TTCTC)$_{19}$TTC(TTCTC)$_2$	108	-1
	Offspring	(TTCTC)$_{18}$TTC(TTCTC)$_2$	103[l]	
14	Parent	(TTCTC)$_{17}$TTC(TTCTC)$_2$	98	+1
	Offspring	(TTCTC)$_{18}$TTC(TTCTC)$_2$	103[l]	
15	Parent	TCCTC(TTCTC)$_5$TTC(TTCTC)$_{28}$TTCT(TTCTC)$_2$TT(TTCTC)$_2$(TC)$_2$	203[j]	-1
	Offspring	TCCTC(TTCTC)$_5$TTC(TTCTC)$_{27}$TTCT(TTCTC)$_2$TT(TTCTC)$_2$(TC)$_2$	198[k]	
16	Parent	TC(TTCTC)$_{29}$TT(TTCTC)$_2$(TC)$_2$	163	+1
	Offspring	TC(TTCTC)$_{30}$TT(TTCTC)$_2$(TC)$_2$	168	
17	Parent	(TTCTC)$_{22}$TT(TTCTC)$_3$TC(TTCTC)(TC)$_2$	138[o]	+1
	Offspring	(TTCTC)$_{23}$TT(TTCTC)$_3$TC(TTCTC)(TC)$_2$	143	
18	Parent	(TTCTC)$_{31}$(TTCCTC)(TTCTC)$_{26}$TT(TTCTC)$_2$(TC)$_2$	307[h]	+1
	Offspring	(TTCTC)$_{32}$(TTCCTC)(TTCTC)$_{26}$TT(TTCTC)$_2$(TC)$_2$	312[i]	
19	Parent	TCCTC(TTCTC)$_{14}$TTC(TTCTC)$_{18}$TTCT(TTCTC)$_2$TT(TTCTC)$_2$(TC)$_2$	198[k]	+1
	Offspring	TCCTC(TTCTC)$_{15}$TTC(TTCTC)$_{18}$TTCT(TTCTC)$_2$TT(TTCTC)$_2$(TC)$_2$	203[j]	
20	Parent	TCCTC(TTCTC)$_{25}$TT(TTCTC)$_2$(TC)$_2$	146[a]	+1
	Offspring	TCCTC(TTCTC)$_{26}$TT(TTCTC)$_2$(TC)$_2$	151[g]	
21	Parent	(TTCTC)$_4$TCCTC(TTCTC)$_{19}$TC(TTCTC)$_{18}$TT(TTCTC)$_2$(TC)$_2$	228[d]	+1
	Offspring	(TTCTC)$_4$TCCTC(TTCTC)$_{19}$TC(TTCTC)$_{17}$TT(TTCTC)$_2$(TC)$_2$	223[c]	
22	Parent	(TTCTC)$_{23}$TC(TTCTC)$_{32}$	277	+1
	Offspring	(TTCTC)$_{24}$TC(TTCTC)$_{32}$	282	
23	Parent	(TTCTC)$_{21}$TTC(TTCTC)$_2$	118[f]	+1
	Offspring	(TTCTC)$_{22}$TTC(TTCTC)$_2$	123[e]	
24	Parent	(TTCTC)$_{20}$TC(TTCTC)$_{34}$	272	+1

	Offspring	(TTCTC)$_{20}$TC(TTCTC)$_{35}$	277	
25	Parent	(TTCTC)$_{24}$TTC(TTCTC)$_2$	133	+1
	Offspring	(TTCTC)$_{25}$TTC(TTCTC)$_2$	138[o]	
26	Parent	TCCTC(TTCTC)$_{20}$TTCTT(TTCTC)$_2$(TC)$_2$	124	-1
	Offspring	TCCTC(TTCTC)$_{19}$TTCTT(TTCTC)$_2$(TC)$_2$	119	

[a] Size homoplasy, identical alleles (Mut1-, Mut4-, Mut11-offspring and Mut20-parent).
[b] Mut5- and Mut6-offspring share parental allele that has mutated, although different outcome of the mutations.
[c] Size homoplasy, heterogenic states (between Mut7-parent and Mut21-offspring).
[d] Size homoplasy, heterogenic states (between Mut7-offspring and Mut21-parent).
[e] Size homoplasy, heterogenic states (between Mut9-parent and Mut23-offspring).
[f] Size homoplasy, heterogenic states (between Mut9-offspring and Mut23-parent).
[g] Size homoplasy, identical alleles (Mut11-parent and Mut20-offspring).
[h] Mut12- and Mut18-parent, the same allele which has mutated and resulted in [i].
[i] Size homoplasy, heterogenic states (see [h]).
[j] Size homoplasy, heterogenic states (between Mut15-parent and Mut19-offspring).
[k] Size homoplasy, heterogenic states (between Mut15-offspring and Mut19-parent).
[l] Size homoplasy, identical alleles (Mut13- and Mut14-offspring).
[o] Size homoplasy, heterogenic states (between Mut17-parent and Mut25offspring)

b) Tree swallow

Mut		**Sequence**	**BP**	**Type**
1	Parent	CTC(TTCTC)$_2$TC(TTCTC)$_{19}$(TC)$_2$(TTCTC)$_3$TT(TTCTC)TC(TTCTC)	143	-2
	Offspring	CTC(TTCTC)$_2$TC(TTCTC)$_{17}$(TC)$_2$(TTCTC)$_3$TT(TTCTC)TC(TTCTC)	133	
2	Parent	TTC(TTCTC)$_2$TC(TTCTC)$_{42}$C	226	+1
	Offspring	TTC(TTCTC)$_2$TC(TTCTC)$_{43}$C	231	
3	Parent	CTC(TTCTC)$_{25}$TT(TTCTC)TCT(TTCTC)TC(TTCTC)	150[m]	-1
	Offspring	CTC(TTCTC)$_{24}$TT(TTCTC)TCT(TTCTC)TC(TTCTC)	145[n]	
4	Parent	CTC(TTCTC)$_{25}$TT(TTCTC)TCT(TTCTC)TC(TTCTC)	150[m]	-1
	Offspring	CTC(TTCTC)$_{24}$TT(TTCTC)TCT(TTCTC)TC(TTCTC)	145[n]	
5	Parent	(TC)$_2$(TTCTC)$_{32}$TTC(TTCTC)TTTC	176	+1
	Offspring	(TC)$_2$(TTCTC)$_{33}$TTC(TTCTC)TTTC	181	
6	Parent	(TC)$_3$TT(TTCTC)$_2$TT(TTCTC)$_{18}$(TCTTCTC)$_6$TC(TTCTC)$_{31}$C	310	+1
	Offspring	(TC)$_3$TT(TTCTC)$_2$TT(TTCTC)$_{18}$(TCTTCTC)$_6$TC(TTCTC)$_{32}$C	315	
7	Parent	C(TTCTC)TT(TTCTC)$_2$TT(TTCTC)$_{17}$(TCTTCTC)$_6$TC(TTCTC)$_{25}$C	275	-1
	Offspring	C(TTCTC)TT(TTCTC)$_2$TT(TTCTC)$_{17}$(TCTTCTC)$_6$TC(TTCTC)$_{24}$C	270	

[m] Allele from same parental individual that mutated.
[n] Same mutation in Mut3 and Mut4, but in two different individuals.

II

Factors affecting germline mutations in a hypervariable microsatellite: a comparative analysis of six species of swallows (Aves: Hirundinidae)

Jarl A. Anmarkrud[a*], Oddmund Kleven[a], Jakob Augustin[b], Kristofer H. Bentz[a], Donald Blomqvist[b], Kim J. Fernie[c], Michael J. L. Magrath[d], Henrik Pärn[a,e], James S. Quinn[f], Raleigh J. Robertson[g], Tibor Szép[h], Scott Tarof[i], Richard H. Wagner[j], and Jan T. Lifjeld[a]

[a] *National Centre for Biosystematics, Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, NO-0318 Oslo, Norway*

[b]*Department of Zoology, University of Gothenburg, Box 463, SE-405 30 Gothenburg, Sweden*

[c] *Ecotoxicology and Wildlife Health, Wildlife and Landscape Science Directorate, Science and Technology Branch, Environment Canada, Burlington, ON, L7R 4A6, Canada*

[d] *Department of Wildlife Conservation and Science, Zoos Victoria, PO Box 74, Parkville, Victoria 3052, Australia*

[e] *Center for Conservation Biology, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*

[f] *Department of biology, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada*

[g] *Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada*

[h] *Institute of Environmental Sciences, College of Nyíregyháza, Nyíregyháza, Sóstói út 31/b, H-4401, Hungary*

[i] *Department of Biology, York University, 4700 Keele Street, Toronto, ON, M3J 1P3,Canada*

[j] *Konrad Lohrenz Institute for Ethology, Austrian Academy of Sciences, Savoyenstrasse 1a, A-1160 Vienna, Austria*

Author for correspondence: * *E-mail: j.a.anmarkrud@nhm.uio.no*

Key words: extrapair paternity, germline mutations, Hirundinidae, *HrU10*, microsatellite evolution, mutation rates, sperm competition

Abstract

Microsatellites mutate frequently by replication slippage. Empirical evidence shows that the probability of such slippage mutations may increase with the length of the repeat region as well as exposure to environmental mutagens, but the mutation rate can also differ between the male and female germline. It has been hypothesized that more intense sexual selection or sperm competition can also lead to elevated mutation rates, but the empirical evidence is inconclusive. Here, we analysed the occurrence of germline slippage mutations in the hypervariable pentanucleotide microsatellite locus *HrU10* across six species of swallow (Aves: Hirundinidae). These species exhibit marked differences in the length range of the microsatellite, as well as differences in the intensity of sperm competition. We found a strong effect of microsatellite length on the probability of mutation, but no residual effect of species or their level of sperm competition when the length effect was accounted for. Neither could we detect any difference in mutation rate between tree swallows (*Tachycineta bicolor*) breeding in Hamilton Harbor, Ontario, an industrial site with previous documentation of elevated mutation rates for minisatellite DNA, and a rural reference population. However, our cross-species analysis revealed two significant patterns of sex differences in *HrU10* germline mutations: 1) mutations in longer alleles occurred typically in the male germline, those in shorter alleles in the female germline, and 2) male germline mutations were more often expansions than contractions, whereas no directional bias was evident in the female germline. These results indicate some fundamental differences in male and female gametogenesis affecting the probability of slippage mutations. Our study also reflects the value of a comparative, multi-species approach for locus-specific mutation analyses, through which a wider range of influential factors can be assessed than in single-species studies.

1. Introduction

Microsatellites are tandemly repeated short nucleotide sequences. Due to their high polymorphism in numbers of repeated units, microsatellites have become popular markers in population genetics with a broad spectrum of applications [1]. The high polymorphism is due to the relatively high mutation rate (10^{-2}-10^{-6} per meiotic event) as compared with

non-repetitive DNA [2], which makes them useful for the study of mutational processes also in non-model species breeding in the wild. Here we investigate how the mutation rates in a highly polymorphic microsatellite, the *HrU10* locus [3], vary within and among six species of swallows (Aves: Hirundinidae), and we try to identify key factors explaining this variation.

The most frequent mechanism for microsatellite mutagenesis is the process of replication slippage [4]. Replication slippage is initiated by proofreading errors during replication [5], and includes a gain or loss of one or more repeated units, compatible with a stepwise mutation model [6]. Several factors may contribute to replication slippage, including motif size [7, 8], chromosomal location [9], G/C-content [10], stabilization pattern within the microsatellite [11], effectiveness of mismatch repair systems [12, 13], potential to create secondary structures [14, 15], environmental conditions [16] and sex [17-20]. Because of the increased probability to introduce slippage events in long tracts of tandemly repeated units, there is generally a positive relationship between microsatellite mutation rate and allele size [17-19, 21-24].

One unresolved issue is whether there is a sex-biased transmission of mutations during meiosis, as contradicting results exist in the literature (e.g.[25, 26]). Most authors have argued for a male driven mutation bias, because males have more germline cell-divisions than females and most mutations originate as replication errors (see [27-29] for reviews). However, this can hardly be true as a general, genome-wide explanation, as there is also evidence of female-biased mutation rates at specific loci, including the *HrU10* [18, 30]. Thus, more information about germline-specific mutation patterns may provide new insights to sex-related factors causing slippage mutations during gametogenesis.

It has also been postulated that intense sexual selection may lead to elevated mutation rates [31]. The logic of this theory is that sexual selection, for example mediated through female mating preferences, may give a high selective premium on mutations that are beneficial (i.e. preferred by females), even when most mutations are not beneficial and will be selected against. This theory may be one solution to the “lek paradox”, which refers to the puzzle of how genetic variation can be maintained in characters subject to sexual selection [32]. Positive relationships between genetic variation and indicators of strong sexual selection have been reported in several studies [33-35]. In particular, a positive relationship between mutation rates in DNA minisatellites and the frequency of extra-pair paternity (EPP) has been claimed for socially monogamous birds [34, 36], but the evidence

has recently been challenged [37]. Whether a similar relationship exists for microsatellites has not yet been tested, despite their extensive use in avian parentage studies.

The nature of microsatellite mutagenesis is difficult to study empirically, especially within natural populations of non-model organisms. Large and correct pedigrees are required in the form of true identification of both genetic parents for a large number of offspring, which may be challenging for both field-work logistics and the availability of markers with sufficiently high mutation rates. On the other hand, parentage analyses have so far been conducted in more than 150 studies of more than 130 bird species [38], and consequently a large body of suitable pedigree data exist from many species. Furthermore, particular microsatellites have high variability and are frequently used as heterologous markers in parentage across a range of species [39], and can thus be specifically targeted for cross-species comparisons. One such marker is the avian microsatellite pentanucleotide *HrU10* [3]. This microsatellite marker is applicable in several bird species, especially within Hirundinidae, and it has been used in several parentage studies [40-45]. *HrU10* has one of the highest mutation rates ever reported and the mutation rates of this marker have also been shown to differ between species [18, 30]. Dawson et al. [46] linked this marker to chromosome 18 in the chicken (*Gallus gallus*) genome. Moreover, Anmarkrud et al. [30] identified an increased *HrU10* mutation rate in tree swallows (*Tachycineta bicolor*) compared to barn swallows (*Hirundo rustica*), and the tree swallow is a species with a substantially higher level of EPP than the barn swallow [40, 41]. Together, this makes *HrU10* a promising candidate marker to test the hypothesis of a positive association between microsatellite mutation rates and the frequency of EPP in birds. Here we expand the previous data set on tree swallows and barn swallows [30] with mutation data from four other hirundine species and an additional population of tree swallows, for which the frequency of EPP is known.

Microsatellite mutation rates might also be affected by the presence of mutagenic substances in the environment, such as radioactive isotopes or polycyclic aromatic hydrocarbons (PAHs). Ellegren et al. [16] documented increased slippage mutation rates at two microsatellite loci in barn swallows breeding in Chernobyl, Ukraine, and exposed to radioactive isotopes released into the local environment in the 1986 accident. Similarly, exposure to PAHs seems to elevate mutation rates in minisatellite DNA in gulls [47] and mice [48] in a polluted industrial area in Hamilton Harbor, Ontario, Canada, though it is unknown whether the same effect applies to microsatellite DNA.

Here, we report the analysis of 100 germline mutations in the *HrU10* marker assembled from six species of swallow (Aves: Hirundinidae), whose populations have previously been subject to parentage analysis. We ask whether the rate of germline mutations differs among the species, and if so, whether this variation can be attributed to variation among species in size range of the microsatellite and/or their level of EPP. We also wanted to test whether the previously reported female mutation bias in tree swallows and barn swallows [30] were upheld in the additional species, and whether we could detect any sex-specific mutation patterns across the entire data set. Finally, we ask whether a study population of tree swallows in the polluted area of Hamilton Harbour, Ontario, showed signs of an elevated mutation rate in this microsatellite as compared to a rural reference population studied previously near Lake Opinicon, Ontario [30].

2. Methods

2.1 General methods

A critical assumption underlying analyses of germline mutations is the correct identification of both genetic parents. We therefore restricted our analysis to data sets containing offspring for which both genetic parents had been identified with a robust parentage system (microsatellites or DNA fingerprints/minisatellites) and for which the population frequency of EPP had been assessed. Our six study species belong to two main clades within the Hirundinidae family: core martins (purple martin *Progne subis*, tree swallow and sand martin *Riparia riparia*) and mud nesters (barn swallow, house martin *Delichon urbicum* and fairy martin *Petrochelidon ariel*) [49]. For three species, the mutation data originate from study populations in North America, viz. purple martin in Pennsylvania [50], barn swallow [30, 41] and tree swallow [40, 51] in Ontario. House martin data were derived from a study population in S Norway [52], the sand martin data from a Hungarian population [53] and the fairy martin data from a population in New South Wales, Australia [54]. Field procedures and paternity analyses for the barn swallow, fairy martin, sand martin and tree swallow are described elsewhere [40, 41, 53, 55]; paternity analyses of the purple martin population are based on S.Tarof and B. Stutchbury, unpublished data. For genotyping at the *HrU10* marker, samples from fairy martin, purple martin and sand martin, were sent to our DNA lab at Natural History Museum in Oslo and genotyped in the same way as previously done in our lab for the other species. All samples that had subject to paternity analysis elsewhere were genotyped with two extra

hypervariable microsatellite markers, *Aar4* and *HrU6*, in addition to *HrU10*, for verification of correct parentage assignment. The mutation data for barn swallow and the rural population of tree swallow were the same as those reported in Anmarkrud et al [30], whereas the parentage analysis and *HrU10* genotyping for the house martin and the tree swallow population in Hamilton Harbour are described in the Appendix. A table with descriptive details of microsatellite variability for all species is also presented there.

For the screening of *HrU10* mutations, we selected a subset of offspring from each population for which both genetic parents were already known from previous parentage analysis and there were no mismatches according to the markers used for parentage verification. Mutations were then defined whenever a *HrU10* allele in the offspring had a mismatch equivalent to the tandemly repeated pentanucleotide unit in any parental alleles. We included only cases of heterozygous offspring, where the offspring's two alleles where verified in the parents. Accordingly, all offspring with potential non-amplifying alleles or apparent homozygosity were excluded from the analyses, as suggested by Ellegren [56]. We assumed the progenitor allele to be the parental allele with the smallest size difference to the mutant allele (cf. Ellegren [56]). All possible mutant alleles were amplified in two independent PCR reactions and analyzed twice for verification of the mutation in both parent and offspring. For each species, we sequenced between 2 and 52 *HrU10* alleles for analysis of the microsatellite structure. The sequence data uncovered the same tandemly repeated core unit $(TTCTC)_n$ in all six species.

2.2 Statistical analysis

Each offspring allele was scored as either mutant or non-mutant, and we recorded its size (bp) and parental origin. We considered each allele as resulting from separate meiotic events, and hence treated them as independent units for statistical analysis. There were in total three cases of possible clustered mutations (see Results), but as a few cases of more than one mutation occurring among offspring from the same parent might also be due to chance, we consider the problem of pseudoreplication to be negligible. The probability of mutations and directionality of mutations were analyzed by logistic regression models with a binominal error distribution and logit link, with respectively mutations (yes/no) or directionality of mutations (expansion/contraction) as a binary response variables, and parental sex, allele size (bp) as predictor variables. The models were fitted using Statistica v7.1 (StatSoft Inc.). Parameter estimates were calculated relative to an intercept representing maternally inherited alleles.

For analyses of the variation in mutation rates among species, a generalised least-squares (GLS) test was applied [57, 58] that controls for phylogenetic non-independence among species. EPP (proportion of young sired by non-social father) and median allele size were included as predictor variables. Briefly, an index of phylogenetic association, λ, is estimated, with values ranging from 0 (indicating phylogenetic independence) to 1 (indicating complete phylogenetic dependence). Likelihood ratio tests were used to compare whether the model with the maximum likelihood value of λ differed from models with λ values of 0 (phylogenetic independence) or 1 (phylogenetic dependence). We followed the phylogeny determined by Sheldon et al. [49] and the analysis was performed in R v2.11.1 [59] using the package APE [60].

3. Results

3.1 Number of mutations and mutation rates

A total of 100 *HrU10* mutations were uncovered among 3604 meiotic events (parent-offspring transmitted alleles) (1802 young), equivalent to an overall mutation rate (μ) of 2.8% (Table 1). The mutation rate varied significantly among the six study species alleles (GLZ; $\chi^2 = 38.01$, $P = < 0.001$) and ranged from 0.6% in house martins to 10.8% in sand martins. The six species also differed in median allele size (Table 1) and there was a positive correlation between mutation rates and median allele sizes at the species level (Fig. 1a). Notably, the species differentiations in allele size seem to reflect their phylogenetic relatedness, as the three species with the longest alleles belong to the core martins and the three species with shortest allele sizes are mud nesters [49]. All mutations involved a gain or loss of one (96 cases) or two (4 cases) tandemly repeated units, consistent to the process of replication slippage [4]. Among the 100 mutations, 60 occurred in a maternally inherited allele and 40 in a paternally inherited allele, indicating an overall maternal bias in mutation rates (Table 2). Three cases of apparently clustered mutations were discovered, that is, siblings with an identical mutant allele from the same parent [61]; One in the barn swallows, one in the purple martins, and one in the tree swallows (Supplementary data 2). Clustered mutations violate one of the key assumptions in population genetic theory, because such mutations can not be considered as independent events. However, as the 100 mutations only revealed siblings with identical mutations in three cases, the concept of clustered mutations was not considered as a major issue, and mutations were treated as independent in the down stream analyses. It should also be

addressed that we identified four cases of mutations that occurred in more than one offspring from the same parent. Yet, these mutations were not consistent with the concept of clustered mutations because the mutations occurred in different inherited alleles. To clarify, a parent may transmit either allele A_1 or allele A_2 to the offspring. In four families, a mutation was observed in allele A_1 in one sibling, and in allele A_2 in another sibling.

3.2 Probability and directions of mutations, allele size and sex

The probability of mutations increased with allele size and was higher in the female than in the male germline (Table 2).). However, the sand martin clearly differed from the other species in having nine of the mutations in the male germline but only two in the female germline (Table 1). The sand martin also had longer alleles than the other species, which accounted for a nearly significant interaction effect between allele size and parental sex. In a model with allele size, parental sex, and their interaction, species was no longer a significant predictor of mutation rate (Table 2).

We also tested for an interaction effect between allele size and parental sex as a function of allele size, analysed separately for mutated (n = 100) and non-mutated (n = 3504) alleles, with the null hypothesis that all mutations, irrespective of allele size, had the same probability of originating from either parent. Using a logistic regression, we found that allele size significantly predicted the parental origin of the mutations (χ^2 = 5.89, P = 0.015), with mutations in longer alleles occurring in the male germline, and those in shorter alleles occurring in the female germline. In contrast, allele size did not predict parental sex in non-mutant alleles (χ^2 = 2.18, P = 0.14). In other words, alleles sizes were similar in the male and female germline, but those that mutated were longer in the male than in the female germline. This analysis does not control for any differences among species, but there is no indication that mutation patterns differ among species when allele size differences are accounted for (cf. Table 2).

Among the mutated alleles, the probability of expansions differed between the sexes (GLZ; χ^2 = 21.54, P < 0.001); expansions were more likely than contractions in males, whereas no pattern in directionality was detected in maternally transmitted mutations (Fig. 2). The direction of the mutation could not be attributed to allele size (GLZ; χ^2 = 0.15.0, P = 0.692).

3.3 Probability of mutations and extra pair paternity

In a species-level comparative analysis, controlling for phylogeny, the species mutation rates were positively related to median allele size (GLS: $n = 6$, $r^2 = 0.32$, $t = 11.88$, $P = 0.001$; $\lambda < 0.01$) but not to levels of extrapair paternity (GLS: $n = 6$, $r^2 = 0.11$, $t = 1.70$, $P = 0.188$; $\lambda < 0.01$) (Fig 1b).

The mutation rate in the sand martins was considerably higher than for any of the other species (Fig. 1b), and may have negated the analysis. Accordingly, we excluded the sand martin and re-analysed the test described above (controlling for shared ancestry). Yet, the species mutation rates were not related to intensity of EPP (GLS: $n = 5$, $r^2 = 0.06$, $t = 0.46$, $P = 0.680$, $\lambda = 0.99$).

3.4 Mutations and PAH exposure

The observed *HrU10* mutation rate for tree swallows in Hamilton Harbor was 5.3 %. The corresponding mutation rate in the rural population was 3.4 % [30]. The difference in mutation rate was far from statistically significant (GLZ; $\chi^2 = 1.60$, $P = 0.210$). When the effect of allele size ($\chi^2 = 17.17$, $P < 0.001$) was included in the model, there was still no between-population effect ($\chi^2 = 1.22$, $P = 0.269$).

4. Discussion

The 100 germline mutations in our data set of genotypes at the *HrU10* microsatellite, assembled from the six swallow species, typically showed either a loss or gain of one repeat unit (5 bp), which is consistent with the mutational process of replication slippage [4]. The probability of mutation clearly increased with allele size. When the effect of allele size was accounted for, there was no residual variation in mutation rate among species that could be attributed to the risk of sperm competition (EPP). Furthermore, there was no elevated mutation rate for tree swallows breeding in Hamilton Harbour exposed to mutagenic PAHs, as compared with birds breeding at the reference site in a presumably unpolluted rural habitat. However, our study revealed two distinct sex differences in mutation patterns of this microsatellite. First, paternally inherited mutant alleles were significantly longer than maternally inherited mutant alleles. Second, the direction of mutations differed between the sexes, with expansions being more likely in the male germline.

It is commonly accepted that microsatellite mutation rates are correlated with allele lengths (e.g. [11, 17-19, 21-24]). Such a pattern may be explained by longer microsatellites being more unstable, with increased probability to introduce errors by the mismatch repair systems during replication. Here we have shown that this pattern also extends to a wider range of allele sizes when analyzed across species. The *HrU10* microsatellite has previously been reported to be highly mutable, with elevated mutation rates in longer alleles in within species comparisons [18, 30]. However, these studies were based on intra-specific analyses only. The observed species-specific differences in allele size, and thereby mutation rates, may reflect evolutionary history and divergence time between the six species. A mutation hypothesis of short microsatellite alleles tending to increase in size and long alleles tending to decrease has been put forward [21, 62]. However, in *HrU10,* we found no effect of allele size on the direction of mutation.

A connection between minisatellite mutation rates and sexual selection, estimated from EPP levels, was proposed by Møller and Cuervo [34]. Recently, Amos [37] criticized this proposal, claiming that Møller and Cuervo's methods for calculating mutation rates were inadequate. In response to this criticism, Møller & Cuervo strongly corroborated their original conclusion, stating that minisatellite mutation rates are positively correlated with sperm competition in birds [36]. In our study, no relationship was observed between the *HrU10* mutation rates and frequency of EPP, but a sample of six species obviously lacks the power to reject this hypothesis. Microsatellites mutate through the process of replication slippage [4] and mutation rates are likely influenced by local chromosomal features such as allele size. Minisatellite mutagenesis is, on the other hand, expected to be induced by double strand breaks following gene conversion-like events [63]. Such differences in mutation mechanisms between micro- and minisatellite DNA, may explain the differences in mutation pattern between the two types of repeated DNA. However, one must keep in mind that our study has examined only on one locus, and there is no way we can infer generality from our results.

The Hamilton Harbour population of tree swallows breeds in an area contaminated with PAHs and other persistent organic pollutants [64]. These chemical compounds are carcinogenic and significantly higher mutation rates have been identified in minisatellites in the herring gull (*Larus argentatus*) and laboratory mice breeding near the same site as the tree swallows in Hamilton Harbour, compared to rural reference populations [47, 48, 65]. However, no such effect was uncovered in the tree swallow population for the *HrU10* microsatellite. Tree swallows are migratory birds that overwinter in the southern USA,

Central America and the Caribbean, whereas herring gulls breeding in Hamilton Harbour overwinter on the Great Lakes and are thereby likely experiencing more exposure to PAHs than the tree swallows. Hence, the relative exposure of the tree swallow to mutagens may be insufficient to alter *HrU10* mutability to detectable levels. However, sperm cells are produced at the beginning of each breeding season [66], and probably after the birds have arrived on their breeding grounds in spring. Spermatogenesis also seems sensitive to mutagens [67]. In that respect, one may expect elevated mutation rates in the paternal germlines, as documented in laboratory mice in the same area [48]. Yet, no sex effect was revealed on mutation rate in the PAH-exposed population compared to the rural reference population (Table 1). The location of the tree swallow colony was south of the area in Hamilton Harbour where herring gulls and mice were exposed to air pollution. Somers et al. [68] reported great differences in the amount of PAH contamination reaching the exposed mice as a function of hours spent downwind of the Hamilton Harbour industrial area. The more southerly location of the tree swallow nest boxes may have resulted in a much lower exposure to airborne contaminants.

Due to a relatively higher number of cell divisions in the male germlines than the female germlines [69], one might expect more germline mutations in paternally than maternally transmitted alleles. An over-all male bias in mutation rate is documented for base substitutions unraveled in avian sex chromosomes [33, 70, 71]. However, we found the opposite pattern for *HrU10*, i.e. mutation rates were higher in female than in male germlines, which is consistent with several previous studies on *HrU10* and other microsatellite loci in natural populations of birds (*Mcyμ8* in superb fairy-wrens *Malurus cyaneus* [17], *HrU10* in barn swallows and in tree swallows [18, 30], *Fn2.14* and *FV2* in the lesser kestrel (*Falco neumanni*) [72]) and other taxa (*Ccon70* in *Camponotus consobrinus* [73], *Ei8* in the olive ridley sea turtle (*Lepidochelys olivacea*) [20]). In contrast, Primmer et al. [24] documented a tendency for higher microsatellite mutation rate in males (*HrU9* locus in barn swallows), and several studies based on multiple microsatellite loci have found higher mutation rates in male germlines in humans (e.g. [19, 74, 75]). Perhaps there is no consistent pattern of sex bias in mutation rates for microsatellites in general, or for particular microsatellites across species. Instead, our results indicate that there was a sex bias related to the length of the microsatellite.

The observed sex heterogeneity in replication slippage patterns suggests specific sex differences during meiosis [76]. Eward et al. [77] documented a striking degree of heterogeneity between male and female gametogenesis with respect to regulation of

replication licensing, the regulative process ensuring precise chromosomal duplication during each cell cycle. Although that study was based on human tissues, important proteins for replication licensing regulation that were found to differ between male and female gametogenesis (e.g. Cdc6, Geminin) are assumed to be present in most eukaryotic cells [78]. One of the replication licensing regulators with sex specific presence in human tissue, Cdc6, functions as a docking platform for the minichromosome maintenance (Mcm2-7) complex [78], which is associated with helicase activity initiating DNA replication [79, 80]. Accordingly, one may speculate that different enzymes are active during DNA replication in male and female meiosis, and the specific enzymatic properties might explain the observed sex dimorphism in replication slippage dynamics.

In conclusion, we have confirmed previous findings of a strong effect of allele size on the likelihood of slippage mutations in a pentanucleotide microsatellite, but also detected new patterns of germline effects on the directionality and allele-size dependency in mutations. We failed to detect any associations with mating system or environmental pollutants, but we will emphasize that such negative results might well be due to a lack of statistical power. Overall, our comparative, multi-species approach has facilitated a broader perspective on microsatellite mutations as compared with single-species studies.

Conflict of interest

We declare that there are no conflicts of interests

Acknowledgements

We thank all who assisted with the field- and lab-work. The research was approved by the Animal Care Committees of Environment Canada and Queen's University, and the Directorate for Nature Management in Norway. This study was supported by grants from the Norwegian Research Council and Natural History Museum, University of Oslo (PhD fellowship JAA), as well as funding through Environment Canada (Ecotoxicology and Wildlife Health).

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version.

References

[1] D.B. Goldstein, C. Schlötterer, Microsatellites: Evolution and applications, Oxford University Press, Oxford, 1999.

[2] Y.C. Li, A.B. Korol, T. Fahima, A. Beiles, E. Nevo, Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review, Mol. Ecol., 11 (2002) 2453-2465.

[3] C.R. Primmer, A.P. Møller, H. Ellegren, New microsatellites from the pied flycatcher *Ficedula hypoleuca* and the swallow *Hirundo rustica* genomes, Hereditas, 124 (1996) 281-283.

[4] G. Levinson, G.A. Gutman, Slipped-strand mispairing: a major mechanism for DNA sequence evolution, Mol. Biol. Evol., 4 (1987) 2032-2021.

[5] M. Strand, T.A. Prolla, R.M. Liskay, T.D. Petes, Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair, Nature, 365 (1993) 274-276.

[6] M. Kimura, T. Ohta, Stepwise mutation model and distribution of allelic frequencies in a finite population, Proc. Natl. Acad. Sci. USA, 75 (1978) 2868-2872.

[7] T.J. Anderson, X.Z. Su, A. Roddam, K.P. Day, Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*, Mol. Ecol., 9 (2000) 1599-1608.

[8] R. Chakraborty, M. Kimmel, D.N. Stivers, L.J. Davison, R. Deka, Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci, Proc. Natl. Acad. Sci. USA, 94 (1997) 1041-1046.

[9] C. Schlötterer, R. Ritter, B. Harr, G. Brem, High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates, Mol. Biol. Evol., 15 (1998) 1269-1274.

[10] K.A. Eckert, G. Yan, S.E. Hile, Mutation rate and specificity analysis of tetranucleotide microsatellite DNA alleles in somatic human cells, Mol. Carcinog., 34 (2002) 140-150.

[11] J.C. Boyer, J.D. Hawk, L. Stefanovic, R.A. Farber, Sequence-dependent effect of interruptions on microsatellite mutation rate in mismatch repair-deficient human cells, Mutat. Res., 640 (2008) 89-96.

[12] B. Harr, J. Todorova, C. Schlötterer, Mismatch repair-driven mutational bias in *D. melanogaster*, Mol. Cell., 10 (2002) 199-205.

[13] J.D. Hawk, L. Stefanovic, J.C. Boyer, T.D. Petes, R.A. Farber, Variation in efficiency of DNA mismatch repair at different sites in the yeast genome, Proc. Natl. Acad. Sci. USA, 102 (2005) 8639-8643.

[14] C.E. Pearson, E.E. Eichler, D. Lorenzetti, S.F. Kramer, H.Y. Zoghbi, D.L. Nelson, R.R. Sinden, Interruptions in the triplet repeats of *SCA1* and *FRAXA* reduce the propensity and complexity of slipped strand DNA (S-DNA) formation, Biochemistry, 37 (1998) 2701-2708.

[15] D.E. Riley, J.S. Jeon, J.N. Krieger, Simple repeat evolution includes dramatic primary sequence changes that conserve folding potential, Biochem. Biophys. Res. Commun., 355 (2007) 619-625.

[16] H. Ellegren, G. Lindgren, C.R. Primmer, A.P. Møller, Fitness loss and germline mutations in barn swallows breeding in Chernobyl, Nature, 389 (1997) 593-596.

[17] N.R. Beck, M.C. Double, A. Cockburn, Microsatellite evolution at two hypervariable loci revealed by extensive avian pedigrees, Mol. Biol. Evol., 20 (2003) 54-61.

[18] J. Brohede, C.R. Primmer, A.P. Møller, H. Ellegren, Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci, Nucleic Acids Res., 30 (2002) 1997-2003.

[19] H. Ellegren, Heterogeneous mutation processes in human microsatellite DNA sequences, Nat. Genet., 24 (2000) 400-402.

[20] W.E. Hoekert, H. Neufeglise, A.D. Schouten, S.B. Menken, Multiple paternity and female-biased mutation at a microsatellite locus in the olive ridley sea turtle (*Lepidochelys olivacea*), Heredity, 89 (2002) 107-113.

[21] Y. Lai, F. Sun, The relationship between microsatellite slippage mutation rate and the number of repeat units, Mol. Biol. Evol., 20 (2003) 2123-2131.

[22] O. Paun, E. Hörandl, Evolution of hypervariable microsatellites in apomictic polyploid lineages of *Ranunculus carpaticola*: directional bias at dinucleotide loci, Genetics, 174 (2006) 387-398.

[23] J.C. Whittaker, R.M. Harbord, N. Boxall, I. Mackay, G. Dawson, R.M. Sibly, Likelihood-based estimation of microsatellite mutation rates, Genetics, 164 (2003) 781-787.

[24] C.R. Primmer, N. Saino, A.P. Møller, H. Ellegren, Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows, Mol. Biol. Evol., 15 (1998) 1047-1054.

[25] H.B. Bohossian, H. Skaletsky, D.C. Page, Unexpectedly similar rates of nucleotide substitution found in male and female hominids, Nature, 406 (2000) 622-625.
[26] K.D. Makova, W.-H. Li, Strong male-driven evolution of DNA sequences in humans and apes, Nature, 416 (2002) 624-626.
[27] H. Ellegren, Characteristics, causes and evolutionary consequences of male-biased mutation, Proc. R. Soc. Lond. B, 274 (2006) 1-10.
[28] L.D. Hurst, H. Ellegren, Sex biases in the mutation rate, Trends in Genetics, 14 (1998) 446-452.
[29] W.-H. Li, S. Yi, K. Makova, Male-driven evolution, Curr. Opin. Genet. Dev., 12 (2002) 650-656.
[30] J. Anmarkrud, O. Kleven, L. Bachmann, J.T. Lifjeld, Microsatellite evolution: Mutations, sequence variation, and homoplasy in the hypervariable avian microsatellite locus *HrU10*, BMC Evol. Biol., 8 (2008) 138.
[31] M. Petrie, G. Roberts, Sexual selection and the evolution of evolvability, Heredity, 98 (2006) 198-205.
[32] M. Kirkpatrick, M.J. Ryan, The evolution of mating preferences and the paradox of the lek, Nature, 350 (1991) 33-38.
[33] A. Bartosch-Harlid, S. Berlin, N.G. Smith, A.P. Møller, H. Ellegren, Life history and the male mutation bias, Evolution, 57 (2003) 2398-2406.
[34] A.P. Møller, J.J. Cuervo, Sexual selection, germline mutation rate and sperm competition, BMC Evol. Biol., 3 (2003) 6.
[35] M. Petrie, C. Doums, A.P. Møller, The degree of extra-pair paternity increases with genetic variability, Proc. Natl. Acad. Sci. USA, 95 (1998) 9390-9395.
[36] A.P. Møller, J. Cuervo, Minisatellite mutation rates increase with extra-pair paternity among birds, BMC Evol. Biol., 9 (2009) 100.
[37] W. Amos, Sexual selection does not influence minisatellite mutation rate, BMC Evol. Biol., 9 (2009).
[38] S.C. Griffith, I.P.F. Owens, K.A. Thuman, Extra pair paternity in birds: a review of interspecific variation and adaptive function, Mol. Ecol., 11 (2002) 2195-2212.
[39] C.R. Primmer, A.P. Møller, H. Ellegren, A wide-range survey of cross-species microsatellite amplification in birds, Mol. Ecol., 5 (1996) 365-378.
[40] K.E. Delmore, O. Kleven, T. Laskemoen, S.A. Crowe, J.T. Lifjeld, R.J. Robertson, Sex allocation and parental quality in tree swallows, Behav. Ecol., 19 (2008) 1243-1249.

[41] O. Kleven, F. Jacobsen, R.J. Robertson, J.T. Lifjeld, Extrapair mating between relatives in the barn swallow: a role for kin selection?, Biol. Lett., 1 (2005) 389-392.

[42] A.P. Møller, J. Brohede, J.J. Cuervo, F. de Lope, C. Primmer, Extrapair paternity in relation to sexual ornamentation, arrival date, and condition in a migratory bird, Behav. Ecol., 14 (2003) 707-712.

[43] C.R. Neuman, R.J. Safran, I.J. Lovette, Male tail streamer length does not predict apparent or genetic reproductive success in North American barn swallows *Hirundo rustica erythrogaster*, J. Avian Biol., 38 (2007) 28-36.

[44] E.L. O'Brien, R.D. Dawson, Context-dependent genetic benefits of extra-pair mate choice in a socially monogamous passerine, Behav. Ecol. Sociobiol., 61 (2007) 775-782.

[45] L.A. Whittingham, P.O. Dunn, Survival of extrapair and within-pair young in tree swallows, Behav. Ecol., 12 (2001) 496-500.

[46] D.A. Dawson, T. Burke, B. Hansson, J. Pandhal, M.C. Hale, G.N. Hinten, J. Slate, A predicted microsatellite map of the passerine genome based on chicken-passerine sequence similarity, Mol. Ecol., 15 (2006) 1299-1320.

[47] C.L. Yauk, J.S. Quinn, Multilocus DNA fingerprinting reveals high rate of heritable genetic mutation in herring gulls nesting in an industrialized urban site, Proc. Natl. Acad. Sci. USA, 93 (1996) 12337-12141.

[48] C.M. Somers, C.L. Yauk, P.A. White, C.L.J. Parfett, J.S. Quinn, Air pollution induces heritable DNA mutations, Proc. Natl. Acad. Sci. USA, 99 (2002) 15904-15907.

[49] F.H. Sheldon, L.A. Whittingham, R.G. Moyle, B. Slikas, D.W. Winkler, Phylogeny of swallows (Aves: Hirundinidae) estimated from nuclear and mitochondrial DNA sequences, Mol. Phylogenet. Evol., 35 (2005) 254-270.

[50] B.J.M. Stutchbury, J.R. Hill III, P.M. Kramer, S.A. Rush, S.A. Tarof, Sex and age-specific annual survival in a Neotropical migratory songbird, the purple martin (*Progne subis*), The Auk, 126 (2009) 278-287.

[51] K.H. Bentz, Germline mutation rate of the microsatellite *HrU10* in tree swallows, *Tachycineta bicolor*, in an area with high PAH pollution, in: The Natural History Museum, University of Oslo, Oslo, 2009, pp. 23.

[52] L.A. Whittingham, J.T. Lifjeld, Extra-pair fertilizations increase the opportunity for sexual selection in the monogamous house martin (*Delichon urbica*), J. Avian Biol., 26 (1995) 283-288.

[53] J. Augustin, D. Blomqvist, T. Szép, Z. Szabó, R. Wagner, No evidence of genetic benefits from extra-pair fertilisations in female sand martins (*Riparia riparia*), J. Ornithol., 148 (2007) 189-198.

[54] M. Magrath, P. Santema, K. Bouwman, D. Brinkhuizen, S. Griffith, N. Langmore, Seasonal decline in reproductive performance varies with colony size in the fairy martin, *Petrochelidon ariel*, Behav. Ecol. Sociobiol., 63 (2009) 661-672.

[55] M.J.L. Magrath, M.A. Elgar, Paternal care declines with increased opportunity for extra-pair matings in fairy martins, Proc. R. Soc. Lond. B, 264 (1997) 1731-1336.

[56] H. Ellegren, Microsatellite mutations in the germline: implications for evolutionary inference, Trends in Genetics, 16 (2000) 551-558.

[57] R.P. Freckleton, P.H. Harvey, M. Pagel, Phylogenetic analysis and comparative data: A test and review of evidence, Am. Nat., 160 (2002) 712-726.

[58] M. Pagel, Inferring the historical patterns of biological evolution, Nature, 401 (1999) 877-884.

[59] R Development Core Team, R: A language and environment for statistical computing, in: R Foundation for Statistical Computing (Ed.), Vienna, Austria, 2009.

[60] E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language, Bioinformatics, 20 (2004) 289-290.

[61] R.C. Woodruff, J.N. Thompson, Have premeiotic clusters of mutation been overlooked in evolutionary theory?, J. Evol. Biol., 5 (1992) 457-464.

[62] Q.Y. Huang, F.H. Xu, H. Shen, H.Y. Deng, Y.J. Liu, Y.Z. Liu, J.L. Li, R.R. Recker, H.W. Deng, Mutation patterns at dinucleotide microsatellite loci in humans, Am. J. Hum. Genet., 70 (2002) 625-634.

[63] A.J. Jeffreys, K. Tamaki, A. MacLeod, D.G. Monckton, D.L. Neil, J.A.L. Armour, Complex gene conversion events in germline mutation at human minisatellites, Nat. Genet., 6 (1994) 136-145.

[64] C.H. Marvin, L. Allan, B.E. McCarry, D.W. Bryant, Chemico/biological investigation of contaminated sediment from the hamilton harbour area of Western Lake Ontario, Environ. Mol. Mutagen., 22 (1993) 61-70.

[65] C.L. Yauk, G.A. Fox, B.E. McCarry, J.S. Quinn, Induced minisatellite germline mutations in herring gulls (*Larus argentatus*) living near steel mills, Mutat. Res., 452 (2000) 211-218.

[66] B. Lofts, R.K. Murton, Reproduction in birds, in: D.S. Farner, J.R. King, K.C. Parkes (Eds.) Avian Biology, Academic Press, London, 1973, pp. 1-107.

[67] A.P. Møller, T.A. Mousseau, C. Lynnn, S. Ostermiller, G. Rudolfsen, Impaired swimming behaviour and morphology of sperm from barn swallows *Hirundo rustica* in Chernobyl, Mut. Res., 650 (2008) 210-216.

[68] C.M. Somers, B.E. McCarry, F. Malek, J.S. Quinn, Reduction of particulate air pollution lowers the risk of heritable mutations in mice, Science, 304 (2004) 1008-1010.

[69] J.F. Crow, The origins, patterns and implications of human spontaneous mutation, Nat. Rev. Genet., 1 (2000) 40-47.

[70] H. Ellegren, A.-K. Fridolfsson, Male-driven evolution of DNA sequences in birds, Nat. Genet., 17 (1997) 182-184.

[71] S. Berlin, M. Brandström, N. Backström, E. Axelsson, N. Smith, H. Ellegren, Substitution rate heterogeneity and the male mutation bias, J. Mol. Evol., 62 (2006) 226-233.

[72] J. Ortego, J.M. Aparicio, P.J. Cordero, G. Calabuig, Characteristics of loci and individuals are associated with germline microsatellite mutation rates in lesser kestrels (*Falco naumanni*), Mutat. Res., 648 (2008) 82-86.

[73] R.H. Crozier, B. Kaufmann, M.E. Carew, Y.C. Crozier, Mutability of microsatellites developed for the ant *Camponotus consobrinus*, Mol. Ecol., 8 (1999) 271-276.

[74] B. Brinkmann, M. Klintschar, F. Neuhuber, J. Huhne, B. Rolf, Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat, Am. J. Hum. Genet., 62 (1998) 1408-1415.

[75] J.L. Weber, C. Wong, Mutation of human short tandem repeats, Hum. Mol. Genet., 2 (1993) 1123-1128.

[76] P.A. Hunt, T.J. Hassold, Sex matters in meiosis, Science, 296 (2002) 2181-2183.

[77] K.L. Eward, E.C. Obermann, S. Shreeram, M. Loddo, T. Fanshawe, C. Williams, H.-I. Jung, A.T. Prevost, J.J. Blow, K. Stoeber, G.H. Williams, DNA replication licensing in somatic and germ cells, J. Cell. Sci., 117 (2004) 5875-5886.

[78] S.P. Bell, A. Dutta, DNA replication in eukaryotic cells, Annu. Rev. Biochem., 71 (2002) 333-374.

[79] K. Labib, J.F.X. Diffley, Is the MCM2-7 complex the eukaryotic DNA replication fork helicase?, Curr. Opin. Genet. Dev., 11 (2001) 64-70.

[80] M. Lei, B. Tye, Initiating DNA synthesis: from recruiting to activating the MCM complex, J. Cell. Sci., 114 (2001) 1447-1454.

Table 1. Details of *HrU10* microsatellite alleles, mutations, male-to-female mutation ratio (α_m), and extrapair paternity for the six study species of swallow (Aves: Hirundinidae).

Species	Common name	Young[a]	Broods[a]	Meiotic events	Mutations	Mutation rate	α_m (♂:♀)	Median allele size[b]	Range	EPP level	References (EPP/mut rates)
Delichon urbicum	House martin	165	58	330	2	0.006	2:0	209	160-298	0.16	This study/This study
Hirundo rustica	Barn swallow	958	223	1916	41	0.021	14:27	231	175-586	0.29	[41]/[30]
Petrochelidon ariel	Fairy martin	173	67	346	5	0.014	2:3	263	196-601	0.20	M. Magrath, unpublished/This study
Progne subis	Purple martin	89	29	178	10	0.056	3:7	310	200-736	0.26	S. Tarof, unpublished./This study
Riparia riparia	Sand martin	51	24	102	11	0.108	9:2	464	322-705	0.19	[53]/This study
Tachycineta bicolor (Hamilton)	Tree swallow	160	53	320	17	0.053	6:11	287	189-592	0.47	This study/[51]
Tachycineta bicolor (QUBS)	Tree swallow	206	56	412	14	0.034	4:10	279	186-605	0.48	[40]/[30]
Over all		1802	510	3604	100	0.028	40:60				

[a] Number of genotyped individuals/broods used for estimation of mutation rates. See Appendix for the total number of genotyped adult individuals for each population.

[b] Adults only.

Table 2. Factors affecting the likelihood of mutation in *HrU10* alleles as shown in a logistic regression model with binominal error distribution (mutation "yes", n = 100 or "no", n = 3504) and logit link function.

Term	*d.f.*	*Estimate*	*St.Error*	χ^2	*P*
Allele size	1	0.0066	<0.0009	53.88	**<0.001**
Parent sex	1	-0.7736	0.3139	6.07	**0.014**
Parent sex × allele size	1	0.0017	0.0009	3.34	0.068
Variable not included in the final model:					
Species	5			4.79	0.443

FIGURE LEGENDS

Figure 1. Relationship between (a) median allele size (bp) and *HrU10* mutation rate per generation, and (b) mutation rate and proportion of EPP in the six study species.

Figure 2. Magnitude and direction of change in the 100 mutations in the *HrU10* microsatellite. Data are presented separately for the male and female germline.

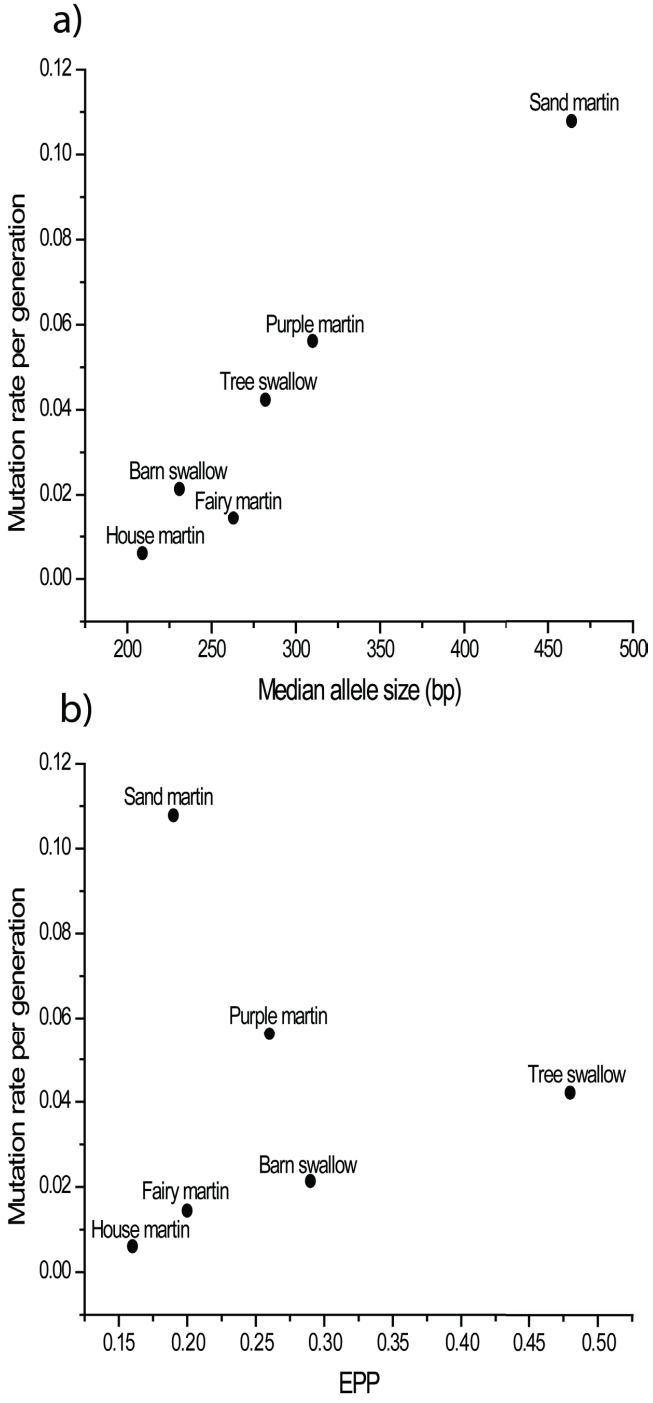
a)
Mutation rate per generation
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Sand martin
Purple martin
Tree swallow
Barn swallow
Fairy martin
House martin
200
250
300
350
400
450
500
Median allele size (bp)
b)
Mutation rate per generation
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Sand martin
Purple martin
Tree swallow
Barn swallow
Fairy martin
House martin
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
EPP

Figure 1.

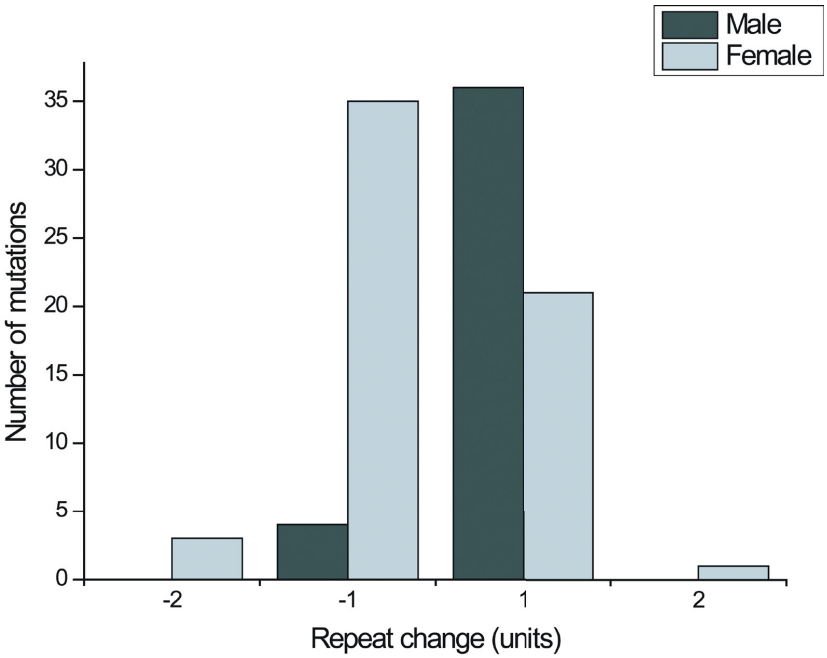
Male
Female
Number of mutations
35
30
25
20
15
10
5
0
-2
-1
1
2
Repeat change (units)

Figure 2.

ELECTRONIC APPENDIX

This is the Electronic Appendix to the article

Factors affecting germline mutations in a hypervariable microsatellite: a comparative analysis of six species of swallows (Aves: Hirundinidae)

by

Jarl A. Anmarkrud, Oddmund Kleven, Jakob Augustin, Kristofer H. Bentz, Donald Blomqvist, Kim J. Fernie, Michael J. L. Magrath, Henrik Pärn, James S. Quinn, Raleigh J. Robertson, Tibor Szép, Scott Tarof, Richard H. Wagner, and Jan T. Lifjeld

<u>This appendix includes the following supplementary data:</u>

Supplementary data 1- **Samples, *HrU10* amplification and genetic analyses**

Supplementary data 2- **Mutation details for every mutation events**

SUPPLEMENTARY DATA 1:

SAMPLES, *HRU10* AMPLIFICATION AND GENETIC ANALYSES

Barn swallow (Hirundo rustica)

The barn swallow *HrU10* data set is the same as in Anmarkrud et al. [1]. A detailed description of the specimens, the markers, their variability and paternity analyses are described by Kleven et al. [2].

Fairy martin (*Petrochelidon ariel*)

Details of field sampling procedures and methodology of paternity analyses are given by Magrath & Elgar [3], and the actual study population is described in Magrath et al. [4]. The data set for *HrU10* mutation analyses consisted of 173 offspring with their genetic parents, and all individuals were genotyped at three microsatellite loci, *HrU10, HrU6* [5] and *Aar4* [6]. The two latter were used to verify correct parentage for both parents. 145 of the offspring had been parentage tested before (M Magrath, unpublished data), whereas the remaining offspring (n=28) were genotyped with an additional triplet of microsatellites at our DNA lab in Oslo using an additional triplet of markers (*Escμ6* [5], *Hir5* [6] and *Hir6* [6]) for correct parentage assignment. See Table 1 for microsatellite details and marker specific exclusion probabilities. The combined non-exclusion probability for second parent (with mother assigned) for the six loci was < 0.00001. The frequency of extrapair paternity for the current study colonies was 0.20 of offspring (M. Magrath, unpublished data), which was somewhat higher than a previous study on another study population (0.14, Magrath & Elgar [3]).

A multiplex polymerase chain reaction (PCR) kit (Qiagen) was employed to amplify *HrU10* (together with *HrU6* and *Aar4*). A 10X primer mix was created consisting of 1.0 μM *Aar4*, 2.0 μM *HrU6* and 2.5 μM *HrU10* fluorescently labelled with FAM, HEX and NED, respectively. The PCR was performed in 6 μL reactions with 1X Qiagen multiplex PCR master mix, 1X primer mix, ~50 ng DNA template and dH_2O to the appropriate volume. The triplet of *Escμ6*, *Hir5* and *Hir6* was amplified in a multiplex under same conditions. Cycle conditions were as recommended for microsatellites by the manufacturer: denaturation in 15 min at 95 °C, followed by 40 cycles with 30 s

denaturation at 94 °C, 90 s annealing at 57 °C, 60 s extension at 72 °C and 30 min final extension at 60 °C. PCR products were genotyped on an ABI3130xl (Applied Biosystems) and the results were analysed in the software GeneMapper v3.0 (Applied Biosystems).

House martin (*Delichon urbicum*)

The House martin samples all originate from a small study colony in Øvre Heimdalen, S. Norway collected over many years 1991-2006). A parentage analysis of samples from the 1991 and 1992 breeding season can be found in Whittingham & Lifjeld [7] and was based on minisatellite DNA fingerprinting. DNA extracts of these samples were genotyped with the same triplet of *HrU10*, *HrU6* and *Aar4*, as described for Fairy martins above. Blood samples from 1994, 1995, 1998, 2001, 2002, 2003, 2005 and 2006 were analysed with this triplet, as well as a second triplet of *Ltr6* [8], *Phtr2* [9] and *Pocc6* [10], for correct parentage assignments. Combined non-exclusion probability was 0.002 for the first triplet and 0.00006 for the combination of all six markers. Microsatellite details are given in Table 1. For the entire data set, extra-pair paternity encompassed 26 of 70 (37.14%) broods and 44 of 277 (15.9 %) the young.

Purple martin (*Progne subis*)

Specimens for analysis originate from a study population in Pennsylvania [11] and paternity analyses for this population has been performed by S. Tarof (unpublished data). *HrU10* was amplified, together with *Aar4* and *HrU6*, from within-pair young and their respective parents only, as described for fairy martin. The combined non-exclusion probability for the triplet of markers was < 0.01. See Table 1 for microsatellite details. The frequency of extrapair for the study population is 0.26 (S. Tarof and B. Stutchbury, unpublished data), which is slightly higher than in another study population in Maryland, USA (0.19; Wagner et al.[12]).

Sand martin (*Riparia riparia*)

The study population in Hungary, field work procedures and paternity analyses (DNA fingerprinting) are described by Augustin et al. [13]. Only within-pair young and their respective parents were selected for analyses with microsatellites. *HrU10* was amplified, together with *Aar4* and *HrU6*, as described for fairy martins. See Table 1 for microsatellite details.

Tree swallow (*Tachycineta bicolor*)

The tree swallow samples originate from two different study populations: one rural nestbox population at Queen's University Biological Station (QUBS), near Chaffey's Lock, Ontario, Canada, studied in 2006, and one nestbox population in Hamilton Harbour, Ontario, Canada studied in 2008. The field work and parentage analyses of the QUBS population are described by Delmore et al. [14] and all *HrU10* mutation data form this population have previously been published in Anmarkrud et al. [1]. The Hamilton population breeds in an industrial area polluted by PAHs and other persistent organic pollutants [15]. PAHs are mutagenic chemical compounds and our intension was to compare *HrU10* mutation rates in Hamilton Harbour to the reference population at QUBS. Paternity analyses and *HrU10* genotyping of the Hamilton samples were performed with six microsatellites; *Aar4*, *HrU10*, *HrU6*, *Ltr6* [8], *Tbi104* [16] and *Pdoμ5* [17]. See Delmore et al. [14] and Stapleton et al. [18] for microsatellite characteristics. The paternity analysis based on the Hamilton samples only, revealed 81.7 % (49/60) of the broods to have extra pair offspring and 46.6% (122/262) of the young were sired by extra pair males. These rates are very similar to the level of extrapair paternity at QUBS (See ref. [14]). Seventeen germline *HrU10* mutations, involving the gain or loss of one or two units, were uncovered among the samples obtained from the Hamilton Harbour population. Combining these mutations with the previously described *HrU10* mutations from the tree swallow population in QUBS [1], provide a total of 31 tree swallow *HrU10* mutations.

Table 1 (next page). Microsatellite details for fairy martin, house martin, purple martin, sand martin and tree swallow. Heterozygosity estimates were performed in Arlequin 3.5.1.2. [19]. *n* refers to number of parental individuals in the *HrU10* mutation rate estimations. *a* refers to number of alleles. H_O an H_E is observed and expected heterozygosity, respectively. Exact test of Hardy-Weinberg equilibrium (HWe p-value) was calculated using 1000000 steps in the Markov chain and 100000 dememorization steps and performed in Arlequin 3.5.1.2. [19]. Polymorphic information content is given as PIC, and average non-exclusion (NE) probability is given for first (1P) and second (2P) parent. Range is given in base pairs (bp). Polymorphic information content and non-exclusion probabilities were calculated in the software Cervus v3.0 [20].

Locus	n[a]	a	H_O	H_E	HWe p-value	PIC	NE-1P	NE-2P	Range (bp)
Fairy martin *(Petrochelidon ariel)*									
Aar4	211	14	0.85	0.83	0.21	0.80	0.51	0.34	107-132
HrU10	195	110	0.76	0.99	0.01	0.98	0.06	0.03	170-496
HrU6	188	167	0.60	0.99	< 0.01	0.99	0.04	0.02	196-601
Escμ6	76	35	0.89	0.96	0.08	0.96	0.16	0.09	108-168
Hir5	77	30	0.90	0.92	0.17	0.91	0.29	0.17	202-278
Hir6	77	51	0.84	0.97	0.01	0.97	0.12	0.07	193-328
House martin *(Delichon urbicum)*									
Aar4	270	20	0.93	0.91	< 0.01	0.90	0.32	0.19	115-149
HrU10	270	68	0.80	0.97	< 0.01	0.97	0.13	0.07	160-298
HrU6	271	36	0.75	0.91	< 0.01	0.91	0.29	0.17	150-218
Ltr6	39	11	0.79	0.79	0.87	0.76	0.57	0.39	185-207
Phtr2	39	18	0.90	0.93	0.44	0.91	0.30	0.18	98-158
Pocc6	33	13	0.76	0.82	0.10	0.54	0.54	0.36	178-211
Purple martin *(Progne subis)*									
Aar4	87	19	0.93	0.89	0.2	0.88	0.37	0.23	111-146
HrU10	84	97	0.90	0.99	< 0.01	0.99	0.06	0.03	200-736
HrU6	84	47	0.87	0.90	0.33	0.90	0.31	0.18	220-369
Sand martin (*Riparia riparia*)									
Aar4	63	10	0.79	0.77	0.45	0.73	0.61	0.43	109-126
HrU10	61	80	0.56	0.99	<0.01	0.98	0.06	0.03	322-721
HrU6	62	62	0.98	0.98	0.18	0.98	0.09	0.05	150-338
Tree swallow (*Tachycineta bicolor*)									
Aar6	120	17	0.89	0.88	0.95	0.87	0.39	0.24	110-142
HrU10	119	113	0.98	0.99	0.11	0.99	0.05	0.03	189-592
HrU6	120	86	0.95	0.94	0.11	0.94	0.20	0.12	150-575
Ltr6	120	11	0.73	0.77	0.56	0.74	0.60	0.42	179-201
Tbi104	120	13	0.76	0.73	0.24	0.70	0.64	0.45	230-255
Pdoμ5	120	8	0.77	0.79	0.95	0.76	0.58	0.40	215-231

[a] Total number of adult individuals genotyped for each marker. Only families with heterozygous *HrU10* genotypes were included in the *HrU10* mutation study (see main article).

References

[1] J. Anmarkrud, O. Kleven, L. Bachmann, J.T. Lifjeld, Microsatellite evolution: Mutations, sequence variation, and homoplasy in the hypervariable avian microsatellite locus *HrU10*, BMC Evol. Biol., 8 (2008) 138.

[2] O. Kleven, F. Jacobsen, R.J. Robertson, J.T. Lifjeld, Extrapair mating between relatives in the barn swallow: a role for kin selection?, Biol. Lett., 1 (2005) 389-392.

[3] M.J.L. Magrath, M.A. Elgar, Paternal care declines with increased opportunity for extra-pair matings in fairy martins, Proc. R. Soc. Lond. B, 264 (1997) 1731-1336.

[4] M. Magrath, P. Santema, K. Bouwman, D. Brinkhuizen, S. Griffith, N. Langmore, Seasonal decline in reproductive performance varies with colony size in the fairy martin, *Petrochelidon ariel*, Behav. Ecol. Sociobiol., 63 (2009) 661-672.

[5] O. Hanotte, C. Zanon, A. Pugh, C. Greig, A. Dixon, T. Burke, Isolation and characterization of microsatellite loci in a passerine bird: the reed bunting *Emberiza schoeniclus*, Mol. Ecol., 3 (1994) 529-530.

[6] O.V. Tsyusko, M.B. Peters, C. Hagen, T.D. Dubberville, T.A. Mousseau, A.P. Møller, T.C. Glenn, Microsatellite markers isolated from barn swallows (*Hirundo rustica*), Mol. Ecol. Notes, 7 (2007) 833-835.

[7] L.A. Whittingham, J.T. Lifjeld, Extra-pair fertilizations increase the opportunity for sexual selection in the monogamous house martin (*Delichon urbica*), J. Avian Biol., 26 (1995) 283-288.

[8] D. McDonald, W. Potts, Cooperative display and relatedness among males in a lek-mating bird, Science, 266 (1994) 1030-1032.

[9] A.-K. Fridolfsson, U.B. Gyllensten, S. Jakobsson, Microsatellite markers for paternity testing in the willow warbler *Phylloscopus trochilus*: High frequency of extra-pair young in an island population, Hereditas, 126 (1997) 127-132.

[10] S. Bensch, T. Price, J. Kohn, Isolation and characterization of microsatellite loci in a Phylloscopus warbler, Mol. Ecol., 6 (1997) 91-92.

[11] B.J.M. Stutchbury, J.R. Hill III, P.M. Kramer, S.A. Rush, S.A. Tarof, Sex and age-specific annual survival in a Neotropical migratory songbird, the purple martin (*Progne subis*), The Auk, 126 (2009) 278-287.

[12] R.H. Wagner, M.D. Schug, E.S. Morton, Condition-dependent control of paternity by female purple martins: implications for coloniality, Behav. Ecol. Sociobiol., 38 (1996) 379-389.

[13] J. Augustin, D. Blomqvist, T. Szép, Z. Szabó, R. Wagner, No evidence of genetic benefits from extra-pair fertilisations in female sand martins (*Riparia riparia*), J. Ornithol., 148 (2007) 189-198.

[14] K.E. Delmore, O. Kleven, T. Laskemoen, S.A. Crowe, J.T. Lifjeld, R.J. Robertson, Sex allocation and parental quality in tree swallows, Behav. Ecol., 19 (2008) 1243-1249.

[15] C.H. Marvin, L. Allan, B.E. McCarry, D.W. Bryant, Chemico/biological investigation of contaminated sediment from the hamilton harbour area of Western Lake Ontario, Environ. Mol. Mutagen., 22 (1993) 61-70.

[16] L.M. Stenzler, Genetic population structure in two bird species with contrasting dispersal behavior: the tree swallow (*Tachycineta bicolor*) and the Florida scrub-jab (*Aphelocoma coerulescens*), in, Cornell University, New York, 2001.

[17] S.C. Griffith, I.R.K. Stewart, D.A. Dawson, I.P.F. Owens, T. Burke, Contrasting levels of extra-pair paternity in mainland and island populations of the house sparrow (Passer domesticus): is there an «island effect»?, Biol. J. Linn. Soc., 68 (1999) 303-316.

[18] M.K. Stapleton, O. Kleven, J.T. Lifjeld, R.J. Robertson, Female tree swallows (*Tachycineta bicolor*) increase offspring heterozygosity through extrapair mating, Behav. Ecol. Sociobiol., 61 (2007) 1725-1733.

[19] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: An integrated software package for population genetics data analysis, Evol. Bioinform. Online, 1 (2005) 47-50.

[20] S.T. Kalinowski, M.L. Taper, T.C. Marshall, Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment, Mol. Ecol., 16 (2007) 1099-1106.

SUPPLEMENTARY DATA 2

MUTATION DETAILS FOR EVERY MUTATION EVENTS

Species	No.	Adult sex	Offspring sex	Parent allele size (bp)	Mutant allele size (bp)	Rep unit diff	Single/ clustered
BARN SWALLOW	1	Female	Male	231	236	1	s
BARN SWALLOW	2	Female	Male	253	248	-1	s
BARN SWALLOW	3	Male	Female	230	235	1	s
BARN SWALLOW	4	Female	Female	210	215	1	s
BARN SWALLOW	5	Female	Female	226	221	-1	s
BARN SWALLOW	6	Female	Female	246	236	-2	s
BARN SWALLOW	7	Female	Male	581	586	1	s
BARN SWALLOW	8	Female	Female	581	576	-1	s
BARN SWALLOW	9	Male	Female	316	321	1	s
BARN SWALLOW	10	Female	Female	251	256	1	c
BARN SWALLOW	11	Female	Female	251	256	1	c
BARN SWALLOW	12	Female	Male	213	208	-1	s
BARN SWALLOW	13	Female	Male	216	211	-1	s
BARN SWALLOW	14	Female	Female	241	236	-1	s
BARN SWALLOW	15	Female	Male	235	230	-1	s
BARN SWALLOW	16	Male	Male	398	403	1	s
BARN SWALLOW	17	Female	Male	198	193	-1	s
BARN SWALLOW	18	Female	Male	188	193	1	s
BARN SWALLOW	19	Female	Male	267	272	1	s
BARN SWALLOW	20	Female	Female	294	289	-1	s

BARN SWALLOW	21	Female	Female	253	258	1	s
BARN SWALLOW	22	Female	Female	251	246	-1	s
BARN SWALLOW	23	Male	Female	233	238	1	s
BARN SWALLOW	24	Female	Male	226	231	1	s
BARN SWALLOW	25	Female	Male	268	263	-1	s
BARN SWALLOW	26	Male	Female	398	403	1	s
BARN SWALLOW	27	Male	Male	289	294	1	s
BARN SWALLOW	28	Female	Male	236	241	1	s
BARN SWALLOW	29	Male	Female	348	353	1	s
BARN SWALLOW	30	Female	Male	321	316	-1	s
BARN SWALLOW	31	Male	Male	368	373	1	s
BARN SWALLOW	32	Male	Female	233	238	1	s
BARN SWALLOW	33	Male	Female	208	213	1	s
BARN SWALLOW	34	Female	Male	241	236	-1	s
BARN SWALLOW	35	Male	Female	213	218	1	s
BARN SWALLOW	36	Male	Female	364	369	1	s
BARN SWALLOW	37	Female	Male	225	230	1	s
BARN SWALLOW	38	Female	Female	214	209	-1	s
BARN SWALLOW	39	Male	Female	251	256	1	s
BARN SWALLOW	40	Female	Male	218	223	1	s
BARN SWALLOW	41	Male	Male	236	241	1	s
FAIRY MARTIN	1	Male	Female	273	278	1	s
FAIRY MARTIN	2	Female	Female	257	252	-1	s
FAIRY MARTIN	3	Female	Female	241	236	-1	s
FAIRY MARTIN	4	Male	Male	345	350	1	s

FAIRY MARTIN	5	Female	Female	232	227	-1	s
HOUSE MARTIN	1	Male	Female	190	195	1	s
HOUSE MARTIN	2	Male	Female	207	212	1	s
PURPLE MARTIN	1	Female	Female	295	300	1	s
PURPLE MARTIN	2	Female	Female	320	325	1	s
PURPLE MARTIN	3	Female	Female	310	305	-1	s
PURPLE MARTIN	4	Female	Male	302	297	-1	c
PURPLE MARTIN	5	Female	Male	302	297	-1	c
PURPLE MARTIN	6	Male	Female	302	424	1	s
PURPLE MARTIN	7	Female	Female	312	307	-1	s
PURPLE MARTIN	8	Male	Female	359	364	1	s
PURPLE MARTIN	9	Female	Female	261	266	1	s
PURPLE MARTIN	10	Male	Male	496	501	1	s
SAND MARTIN	1	Male	Male	507	502	-1	s
SAND MARTIN	2	Male	Female	456	461	1	s
SAND MARTIN	3	Male	Female	593	588	-1	s
SAND MARTIN	4	Female	Female	455	450	-1	s
SAND MARTIN	5	Male	Male	446	451	1	s
SAND MARTIN	6	Male	Male	438	443	1	s
SAND MARTIN	7	Male	Female	436	441	1	s
SAND MARTIN	8	Male	Male	419	424	1	s
SAND MARTIN	9	Male	Female	395	400	1	s
SAND MARTIN	10	Female	Female	534	529	-1	s
SAND MARTIN	11	Male	Male	549	554	1	s
TREE SWALLOW	1	Female	Female	232	222	−2	s

TREE SWALLOW	2	Male	Male	505	510	+1	s
TREE SWALLOW	3	Female	Male	388	398	+2	s
TREE SWALLOW	4	Male	Male	318	323	+1	s
TREE SWALLOW	5	Female	Female	281	276	−1	s
TREE SWALLOW	6	Male	Male	338	343	+1	s
TREE SWALLOW	7	Male	Male	284	289	+1	s
TREE SWALLOW	8	Female	Male	238	233	−1	s
TREE SWALLOW	9	Female	Female	270	275	+1	s
TREE SWALLOW	10	Female	Male	511	506	−1	s
TREE SWALLOW	11	Female	Male	265	270	+1	s
TREE SWALLOW	12	Female	Female	523	518	−1	s
TREE SWALLOW	13	Female	Male	400	405	+1	s
TREE SWALLOW	14	Female	Female	365	360	−1	s
TREE SWALLOW	15	Male	Female	282	287	+1	s
TREE SWALLOW	16	Male	Female	575	580	+1	s
TREE SWALLOW	17	Female	Female	324	319	−1	s
TREE SWALLOW	18	Male	Female	348	343	−1	s
TREE SWALLOW	19	Female	Male	462	457	−1	s
TREE SWALLOW	20	Female	Male	303	298	−1	s
TREE SWALLOW	21	Female	Female	286	291	+1	c
TREE SWALLOW	22	Female	Male	286	291	+1	c
TREE SWALLOW	23	Female	Male	251	246	−1	s
TREE SWALLOW	24	Male	Female	484	489	+1	s
TREE SWALLOW	25	Female	Male	284	279	−1	s
TREE SWALLOW	26	Male	Male	257	252	−1	s

TREE SWALLOW	27	Male	Male	302	307	+1	s
TREE SWALLOW	28	Female	Female	354	349	−1	s
TREE SWALLOW	29	Female	Female	293	298	+1	s
TREE SWALLOW	30	Female	Male	568	563	−1	s
TREE SWALLOW	31	Female	Male	592	580	−2	s

III

IV